

RD 9/29/99
Result 14 static OK Page 1
but no 3A conf. See .

1	1491	100.0	1491	22	AAE77666	Human
2	1491	100.0	2272	22	AAE77667	Human
3	1487.8	99.8	4877	22	FAE60871	Human
4	1487.8	99.8	4877	22	FAE60872	Human
5	1487.8	99.8	4877	22	AAE60883	Human
6	1486.2	99.9	2400	22	AAE77264	Human
7	1474.8	98.9	2403	22	AAE77265	Human
8	1474.8	97.4	2412	22	AAE77676	Human
9	1452.6	93.5	1669	22	AAE77675	Human

ALIGNMENTS

KW Human, protease-inhibitor like protein; NHP; novel human protein;
KW trypsin inhibitor; gene therapy; drug screening; disease diagnosis
KW signal transduction; ss.

OS Homo sapiens

PN WO200121651-A2

PD 29-MAR-2001.

22-SEP-2000; 2000WO-US26048

PR 24-SEP-1999; 99US-0156101

PA (LEXI-) LEXICON GENETICS INC.
XX

PI	Donoho G,	Turner CA,	Wattler F,	Nellis M,	Friedrich G,
PI	Zambrowicz B	Sands AT:			

PI Zambrowicz B, Sands AT;

XX
XX
WDT: 2001-266064/27

DR WP1; 2001-200004,
DB D-PSNB: AAB7A445

XX

PT Novel isolated human protease inhibitor-like polynucleotide useful in
 therapeutic, diagnostic and pharmacogenic applications
 Claim 1: Page 26; 29pp; English.

CC The present invention provides the protein and coding sequences of the
 CC novel human protease-inhibitor like protein NHP. This shows homology to
 CC mammalian trypsin inhibitors. The sequences are useful in disease
 CC diagnosis and treatment, particularly of diseases associated with signal
 CC transduction. The present sequence is one version of the NHP cDNA.

XX Sequence 1491 BP; 368 A; 409 C; 426 G; 288 T; 0 other:

Query Match 100.0%; Score 1491; DB 22; Length 1491;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1491; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 atgagctgctctctggtggtgtacacccctggggctgctgttctctgttgcggatcc 60
DB 1 atgagctgctctctggtggtgtacacccctggggctgctgttctctgttgcggatcc 60
QY 61 caaggtactctctgcccacgtctactctcttaagagagctgctcaagaataccagcac 120
DB 61 caaggtactctctgcccacgtctactctcttaagagagctgctcaagaataccagcac 120
QY 121 aacgagctctctctggtggtgtacacccctggggctgctgttctctgttgcggatcc 180
DB 121 aacgagctctctctggtggtgtacacccctggggctgctgttctctgttgcggatcc 180
QY 181 atgctgcacaaagctctcgggagctgagctcagctcagctcacaatgagatcatg 240
DB 181 atgctgcacaaagctctcgggagctgagctcagctcagctcacaatgagatcatg 240
QY 241 aactggggtgacgaactgagagaatctgctgagcgtggtggcgaatcagatctggag 300
DB 241 aactggggtgacgaactgagagaatctgctgagcgtggtggcgaatcagatctggag 300
QY 301 caggggcccacacatctctgctgttccatcgcggagagaacccgtggccttactctgggag 360
DB 301 caggggcccacacatctctgctgttccatcgcggagagaacccgtggccttactctgggag 360
QY 361 tatgctctcggggttccatctgagctcctggtatgacgaggtgaaaggactaacctac 420
DB 361 tatgctctcggggttccatctgagctcctggtatgacgaggtgaaaggactaacctac 420
QY 421 cccctaccgagagtgacacccctgtgtccagagaggtgctcgggctactgtgacg 480
DB 421 cccctaccgagagtgacacccctgtgtccagagaggtgctcgggctactgtgacg 480
QY 481 caactacacagatagtttggtccacacacacacacacacacacacacacacacacac 540
DB 481 caactacacagatagtttggtccacacacacacacacacacacacacacacacacac 540
QY 541 cggagagtgactgtctggtgagaggttggtgagagcgggtacttctgtctgcaattat 600
DB 541 cggagagtgactgtctggtgagaggttggtgagagcgggtacttctgtctgcaattat 600
QY 601 tctccaaaggagagctggtgagagagcccttacaagaatgcccggcctgtctgag 660
DB 601 tctccaaaggagagctggtgagagagcccttacaagaatgcccggcctgtctgag 660
QY 661 tgcacacacagctatgagagcagctcagagaacacttgttaccagagaagaactac 720
DB 661 tgcacacacagctatgagagcagctcagagaacacttgttaccagagaagaactac 720
QY 721 actccaaacactgaaacgagagatgagatgagtgagaaagcgtccatctctgaaga 780
DB 721 actccaaacactgaaacgagagatgagatgagtgagtgagaaagcgtccatctctgaaga 780
QY 781 aaccatgttggctccacacgaggtgagtgagagccacacacacacacacacacacac 840
DB 781 aaccatgttggctccacacgaggtgagtgagagccacacacacacacacacacacac 840

```

```

QY 841 gtcaactacagaccccaagctgtcagatgtgacaccaaagtgaagagacaggtgcaagg 900
DB 841 gtcaactacagaccccaagctgtcagatgtgacaccaaagtgaagagacaggtgcaagg 900
QY 901 tccacgttaccaggtaccagctgcccagcagcgtgctgtaacacaaaggagagatctt 960
DB 901 tccacgttaccaggtaccagctgcccagcagcgtgctgtaacacaaaggagagatctt 960
QY 961 ggaagctgttctatgaagctcgtctagacatagccgcgcgcacacacacacacac 1020
DB 961 ggaagctgttctatgaagctcgtctagacatagccgcgcgcacacacacacacac 1020
QY 1021 ctgagtagaaggagagcgcgtggtgagatataccagagacgagaggtccctctctgtg 1080
DB 1021 ctgagtagaaggagagcgcgtggtgagatataccagagacgagaggtccctctctgtg 1080
QY 1081 aagcttgagagacagcgcgtgagctccctcagacaaataacacacacacacacacac 1140
DB 1081 aagcttgagagacagcgcgtgagctccctcagacaaataacacacacacacacacac 1140
QY 1141 gtttcaaaagtgaaagtgtcagagatttggactgtacacgacggtgtcagctgtgccg 1200
DB 1141 gtttcaaaagtgaaagtgtcagagatttggactgtacacgacggtgtcagctgtgccg 1200
QY 1201 ttgaaagacagacacacacacacacacacacacacacacacacacacacacacacac 1260
DB 1201 ttgaaagacagacacacacacacacacacacacacacacacacacacacacacacac 1260
QY 1261 ccttctactctggtcgcgtgtgttggaaacacacacacacacacacacacacacac 1320
DB 1261 ccttctactctggtcgcgtgtgttggaaacacacacacacacacacacacacacac 1320
QY 1321 aagacagcgtgtgacgcgggaggtcatcagacacacacacacacacacacacacacac 1380
DB 1321 aagacagcgtgtgacgcgggaggtcatcagacacacacacacacacacacacacacac 1380
QY 1381 cccgtgataaaaaaagaacacacacacacacacacacacacacacacacacacacac 1440
DB 1381 cccgtgataaaaaaagaacacacacacacacacacacacacacacacacacacacac 1440
QY 1441 ctgggagctcctcgggagtgagagccttcggatcttctgtcgaagcag 1491
DB 1441 ctgggagctcctcgggagtgagagccttcggatcttctgtcgaagcag 1491

```

RESULT 2
 AAF77687 standard; cDNA; 2272 BP.
 XX AAF77687;
 DT 29-MAY-2001 (first entry)
 XX Human protease-inhibitor like protein coding sequence #2.
 DE Human; protease-inhibitor like protein; NHP; novel human protein;
 KW trypsin inhibitor; gene therapy; drug screening; disease diagnosis;
 KW signal transduction; ss.
 OS Homo sapiens.
 XX Homo sapiens.
 PN WO200121651-A2.
 PD 29-MAR-2001.
 XX 22-SEP-2000; 2000WO-US26048.
 XX 24-SEP-1999; 99US-0156101.
 XX (LEXI-) LEXICON GENETICS INC.
 PA Donoho G, Turner CA, Wattler F, Nehls M, Friedrich G;
 XX

SEQ ID NO: 1
Ac NO: AAD17766
Database: N-GeneSeq-032802

```

421 cccaccgagcgagtgcaacccctggtgtccagagagtgctcgggcctatgtgacg 480
    |||||||
626 cccaccgagcgagtgcaacccctggtgtccagagagtgctcgggcctatgtgacg 685
    |||||||
481 cactacacacagatgtgttggtccacacacacacacacacacacacacacacac 540
    |||||||
686 cactacacacagatgtgttggtccacacacacacacacacacacacacacacac 745
    |||||||
541 cggagagatgactgtcttggtggagagagtttggtgagaaacggtctacttctgtcaattat 600
    |||||||
746 cggagagatgactgtcttggtggagagagtttggtgagaaacggtctacttctgtcaattat 805
    |||||||
601 tctccaaagggaacttgatctggagaaagccctccaaagatctggccgctgtctgag 660
    |||||||
806 tctccaaagggaacttgatctggagaaagccctccaaagatctggccgctgtctgag 865
    |||||||
661 tggccaccacgactatgagagagagctgtcagaaacacattgtgttaccgagaaacacac 720
    |||||||
866 tggccaccacgactatgagagagagctgtcagaaacacattgtgttaccgagaaacacac 925
    |||||||
721 actccaaacacgagaaacgagagagatgagatgagatgagatgagatgagatgagatgagat 780
    |||||||
926 actccaaacacgagaaacgagagagatgagatgagatgagatgagatgagatgagatgagat 985
    |||||||
781 aacacatgttggctccacacgagagtgatgagacccacacacacacacacacacacacac 840
    |||||||
986 aacacatgttggctccacacgagagtgatgagacccacacacacacacacacacacacac 1045
    |||||||
841 gtcacacacacacacacacacacacacacacacacacacacacacacacacacacacacacac 900
    |||||||
1046 gtcacacacacacacacacacacacacacacacacacacacacacacacacacacacacacac 1105
    |||||||
901 tccacagtgtaacacagtgtaacacagtgtaacacagtgtaacacagtgtaacacagtgtaacac 960
    |||||||
1106 tccacagtgtaacacagtgtaacacagtgtaacacagtgtaacacagtgtaacacagtgtaacac 1165
    |||||||
961 ggaacatgttggctccacacgagagtgatgagacccacacacacacacacacacacacac 1020
    |||||||
1166 ggaacatgttggctccacacgagagtgatgagacccacacacacacacacacacacacac 1225
    |||||||
1021 ctgagtgacacagagagagcctgtgtgataccacacacacacacacacacacacacacacac 1080
    |||||||
1226 ctgagtgacacagagagagcctgtgtgataccacacacacacacacacacacacacacacac 1285
    |||||||
1081 aagtcgtgagagacacagagcctgtgtgataccacacacacacacacacacacacacacacac 1140
    |||||||
1286 aagtcgtgagagacacagagcctgtgtgataccacacacacacacacacacacacacacacac 1345
    |||||||
1141 gtcgtcaaaagtgaagagtgacagatctgtgacacacacacacacacacacacacacacacac 1200
    |||||||
1346 gtcgtcaaaagtgaagagtgacagatctgtgacacacacacacacacacacacacacacacac 1405
    |||||||
1201 ttgtgaaagccacacacacacacacacacacacacacacacacacacacacacacacacacac 1260
    |||||||
1406 ttgtgaaagccacacacacacacacacacacacacacacacacacacacacacacacacacac 1465
    |||||||
1261 ccttcctactggtggtcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1320
    |||||||
1466 ccttcctactggtggtcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1525
    |||||||
1321 aagacagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1380
    |||||||
1526 aagacagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1585
    |||||||
1381 cccgtgtgataaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaa 1440
    |||||||
1586 cccgtgtgataaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaa 1645
    |||||||
1441 ctggggagactcctcgagagagaaagccttcgagatcttgcctgtgacagcag 1491
    |||||||
1646 ctggggagactcctcgagagagaaagccttcgagatcttgcctgtgacagcag 1696
    |||||||

```

```

RESULT 7
AAD17766
ID AAD17766 standard; DNA; 2403 BP.
XX
AC AAD17766;
XX
DE 10-DEC-2001 (first entry)
XX
DE Human novel trypsin inhibitor-like protein, NOV-4d encoding DNA.
XX
XX Human; NOV-X protein; KIAA1233-like protein; STE20-like protein; tumour;
XX trypsin inhibitor-like protein; gene therapy; haematopoietic; illness;
XX immunological disorder; neurodegenerative disorder; Alzheimer's disease;
XX Parkinson's disease; immunomodulatory; pharmacogenomic; haemostatic;
XX human immunodeficiency virus; HIV; fertility disorder; neuroprotective;
XX cytosolic; nucleolar; anti-fertility; cancer; NOV-4d protein; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 206..1702
FT /tag= a
FT /product= "Human novel trypsin inhibitor-like protein,
FT NOV-4d"
XX
PN WO200162928-A2.
XX
PD 30-ANG-2001.
XX
XX 26-FEB-2001; 2001WO-US06151.
XX
XX 25-FEB-2000; 2000US-0184951.
XX 28-FEB-2000; 2000US-0185548.
XX 01-MAR-2000; 2000US-0185967.
XX 18-APR-2000; 2000US-0197723.
XX 27-APR-2000; 2000US-0199957.
XX 23-FEB-2001; 2001US-0789390.
XX
PA (CURA-) CURAGEN CORP.
XX
PT Vernet CAM, Fernandes E, Shinkets RA, Macdougall J, Spaeferna SK;
XX WPI; 2001-582051/65.
XX DR P-PDB: AAE17766.
XX
XX New isolated KIAA1233-like, STE20-like, or trypsin inhibitor-like
XX polypeptide for diagnosing and treating pathological disorders, such as
XX Parkinson's disease and for use in pharmacogenomics -
XX
XX Claim 9; Page 87-88; 189pp; English.
XX
XX The invention relates to novel human polypeptides referred as NOV-X
XX and their corresponding nucleic acid sequences. NOV-X collectively
XX include NOV-1, NOV-2a and NOV-2b which are novel KIAA1233-like
XX polypeptides, NOV-3a, NOV-3b, NOV-3c and NOV-3d which are novel
XX STE20-like polypeptides and NOV-4a, NOV-4b, NOV-4c, NOV-4d and NOV-4e
XX which are novel trypsin inhibitor-like polypeptides. NOV-X is used to
XX identify a potential therapeutic agent that can modulate its activity
XX and can be used for treating a pathology related to aberrant expression
XX or aberrant physiological interactions of NOV-X. NOV-X or its DNA is
XX used to determine the presence or predisposition to a disease associated
XX with altered levels of NOV-X. NOV-X, its DNA and its antibody are used
XX to treat or prevent a pathology associated with NOV-X. The pathological
XX states that can be treated or prevented are haematopoietic, cancer,
XX immunological, tumour, neurodegenerative (e.g. Alzheimer's and
XX Parkinson's disease), human immunodeficiency virus (HIV) illness and
XX fertility disorders. NOV-X and its DNA are used in pharmacogenomics for
XX predictive medicine. NOV-X DNA is used in gene therapy. The present
XX sequence is a DNA encoding human novel trypsin inhibitor-like protein,
XX NOV-4d.
XX
SQ Sequence 2403 BP; 558 A; 646 C; 685 G; 514 T; 0 other;

```


Db	1706	cag	1708	
XX	RESULT	9		
XX	AADI17765			
XX	AADI17765	standard; DNA; 1669 BP.		
XX	AADI17765;			
XX	10-DEC-2001	(first entry)		
DE	Human novel trypsin inhibitor-like protein, NOV-4c encoding DNA.			
XX	Human: NOV-X protein; KIA1233-like protein; STE20-like protein; tumour;			
KW	trypsin inhibitor-like protein; gene therapy; haematopoietic; illness;			
KW	immunological disorder; neurodegenerative disorder; Alzheimer's disease;			
KW	Parkinson's disease; immunomodulatory; pharmacogenomic; hemostatic;			
KW	human immunodeficiency virus; HIV; fertility disorder; neuroprotective;			
XX	cytostatic; neutrotropic; anti-infertility; cancer; NOV-4c protein; ds.			
XX				
OS	Homo sapiens.			
XX				
FH	Key	Location/Qualifiers		
FT	CDS	205..1638		
FT		/*tag= a		
FT		/product= "Human novel trypsin inhibitor-like protein,		
FT	sig_peptide	NOV-4c"		
FT		205..270		
FT		/*tag= b		
FT	mat_peptide	271..1635		
FT		/*tag= c		
FT		/product= "Human mature novel trypsin inhibitor-like		
FT		protein, NOV-4c"		
XX				
PN	WO200162928-A2.			
XX				
PD	30-AUG-2001.			
XX				
PF	26-FEB-2001; 2001WO-US06151.			
XX				
PR	25-FEB-2000; 2000US-0184951.			
PR	28-FEB-2000; 2000US-0185548.			
PR	01-MAR-2000; 2000US-0185967.			
PR	18-APR-2000; 2000US-0197723.			
PR	27-APR-2000; 2000US-0199957.			
PR	23-FEB-2001; 2001US-0789390.			
XX				
PA	(CURA-) CURAGEN CORP.			
PI	Vernet CAM, Fernandes E, Shimkets RA, Macdougall J, Spaderna SK;			
XX	WPI; 2001-582051/65.			
DR	P-PSDB; AAE10617.			
PT	New isolated KIA1233-like, STE20-like, or trypsin inhibitor-like			
XX	polypeptide for diagnosing and treating pathological disorders, such as			
PT	Parkinson's disease and for use in pharmacogenomics -			
XX				
PS	Claim 9; Page 85; 189pp; English.			
XX				
CC	The invention relates to novel human polypeptides referred as NOV-X			
CC	and their corresponding nucleic acid sequences. NOV-X collectively			
CC	include NOV-1, NOV-2a and NOV-2b which are novel KIA1233-like			
CC	polypeptides, NOV-3a, NOV-3b, NOV-3c and NOV-3d which are novel			
CC	STE20-like polypeptides and NOV-4a, NOV-4b, NOV-4c, NOV-4d and NOV-4e			
CC	which are novel trypsin inhibitor-like polypeptides. NOV-X is used to			
CC	identify a potential therapeutic agent that can modulate its activity			
CC	and can be used for treating a pathology related to aberrant expression			
CC	or aberrant physiological interactions of NOV-X. NOV-X or its DNA is			
CC	used to determine the presence or predisposition to a disease associated			
CC	with altered levels of NOV-X. NOV-X, its DNA and its antibody are used			
CC	to treat or prevent a pathology associated with NOV-X. The pathological			

Query Match	93.5%	Score 1393.8	DB 22	Length 1659
Best Local Similarity	99.2%	Pred. No. 0		
Matches 1401	Conservative 0	Mismatches 12	Indels 0	Gaps 0
OY 1 atgaactgcgtccttggtggtgtatcacccttgggagcgtctgtctctgtcgtgagacc 60				
Db 205 atgaactgcgtccttggtggtgtatcacccttgggagcgtctgtctctgtcgtgagacc 264				
OY 61 caaagctacactcctctcccaacgctacactctctatagaggagctctgctacgaataacagcac 120				
Db 265 caaagctacactcctctcccaacgctacactctctatagaggagctctgctacgaataacagcac 324				
OY 121 aacgagctcactcccgaggtcccgagagccatccccaaggagaaagagagagatcttc 180				
Db 325 aacgagctcactcccgaggtcccgagagccatccccaaggagaaagagagagatcttc 384				
OY 181 atgcgtcacaacaagacttcgggccaaggtgagacgctcagcgtcccaactcaatgtagatcag 240				
Db 385 atgcgtcacaacaagacttcgggccaaggtgagacgctcagcgtcccaactcaatgtagatcag 444				
OY 241 acctgggagatgacgaactggaagaagctctgctcagcgttggccagtcagtgtaccttggag 300				
Db 445 acctgggagatgacgaactggaagaagctctgctcagcgttggccagtcagtgtaccttggag 504				
OY 301 caccggggccccaacagactgtgtgttccatctcgggccaagaaacttggcgctacatggggcag 360				
Db 505 caccggggccccaacagactgtgtgttccatctcgggccaagaaacttggcgctacatggggcag 564				
OY 361 tatcgtctctccggggtttccatgttgacgttccctgtgatacgaagggtgaaggactacactac 420				
Db 565 tatcgtctctccggggtttccatgttgacgttccctgtgatacgaagggtgaaggactacactac 624				
OY 421 ccctaccgagcgagatgcaacccttgggttccaaggaggttgcttcgggggcttatgtgcaag 480				
Db 625 ccctaccgagcgagatgcaacccttgggttccaaggaggttgcttcgggggcttatgtgcaag 684				
OY 481 caactacacacatatagtttgggccaacacacaagaatcgtgtgtctcttgacaacctgc 540				
Db 685 caactacacacatatagtttgggccaacacacaagaatcgtgtgtctcttgacaacctgc 744				
OY 541 cggagaatgactgtcttgggagagaagtttgggagaaacgcggtctacttgtctgcaattat 600				
Db 745 cggagaatgactgtcttgggagagaagtttgggagaaacgcggtctacttgtctgcaattat 804				
OY 601 tctcccaagggaactggaacttgggagaagcccttacaagaatggcgcgccctgacttag 660				
Db 805 tctcccaagggaactggaacttgggagaagcccttacaagaatggcgcgccctgacttag 864				
OY 661 tggccaccagatgatagaggagcgtgcaagaaacacttgttatacggaaagaagaactac 720				
Db 865 tggccaccagatgatagaggagcgtgcaagaaacacttgttatacggaaagaagaactac 924				
OY 721 accccaagaactgaaacggagagatgaaattgaaattgaaacgggtcccttcccttgaaaga 780				
Db 925 accccaagaactgaaacggagagatgaaattgaaattgaaacgggtcccttcccttgaaaga 984				
OY 781 aaccatgcttgggtcccaacggaggtgatatgagcccaacgaagcccaagaanaacctctgg 840				
Db 985 aaccatgcttgggtcccaacggaggtgatatgagcccaacgaagcccaagaanaacctctgg 1044				
OY 841 gtaactacatgacccaagtcttcaagatgtgataccaagaatgaaggaaacggtgcaagaag 900				

Db	1045	gtcaacttaacatgaccacaagctcgtcttatgttgaaacccaagaatgaaggaacaggtgtgcaaaagg	1104
QY	901	tccacgtgtgtaacacaggtatccacgtgtcccaagcagctctccctgaacccaagaagcggaagacttt	960
Db	1105	tccacgtgtgtaacacaggtatccacgtgtcccaagcagctctccctgaacccaagaagcggaagacttt	1164
QY	961	ggaagctctgtctctctgaagaactcgtctctgaacatctgctgcgcgcgcgcacatccacatacgggatac	1020
Db	1165	ggaactctgtctctctgaagaactcgtctctgaacatctgctgcgcgcgcgcacatccacatacgggatac	1224
QY	1021	ctggatgtacaaaggagagcctcgtgtgatatcaaccaagaaacgggaaggtccctcttcgtg	1080
Db	1225	ctggatgtacaaaggagagcctcgtgtgatatcaaccaagaaacgggaaggtccctcttcgtg	1284
QY	1081	aagcgtctagaagacacgcgcgtgcagctctccctcagaacaatacaaacctctccagctctatctacg	1140
Db	1285	aagcgtctagaagacacgcgcgtgcagctctccctcagaacaatacaaacctctccagctctatctacg	1344
QY	1141	gtgtgcacaagaatgaaagtgtcagaagatttggactctgtacaacgaccgltgtctagctgtgcccg	1200
Db	1345	gtgtgcacaagaatgaaagtgtcagaagatttggactctgtacaacgaccgltgtctagctgtgcccg	1404
QY	1201	tttcaaaagcgcaacaactcaactcgtcccaagaataccattgtccggcgacatctgnaaagaacga	1260
Db	1405	tttcaaaagcgcaacaactcaactcgtcccaagaataccattgtccggcgacatctgnaaagaacga	1464
QY	1261	cctctctactcgtggtcccggtgttggaaaccaactatgacagatacctctcaagaacatctgc	1320
Db	1465	cctctctactcgtggtcccggtgttggaaaccaactatgacagatacctctcaagaacatctgc	1524
QY	1321	aagaacacctctgacacgcgggagatcatcaacacacgagaatgtgggggtgtgacgttgaaagctgatg	1380
Db	1525	aagaacacctctgacacgcgggagatcatcaacacacgagaatgtgggggtgtgacgttgaaagctgatg	1584
QY	1381	cccggtgtatataaagaagaactcagctgtggctcg	1413
Db	1585	cccggtgtatataaagaagaactcagctgtggctcg	1617

XX	RESULT	10
AD17763	AD17763	
ID	AD17763	standard; DNA; 2305 BP.
XX		
AC	AD17763;	
XX		
DT	10-DEC-2001	(first entry)
XX		
DE	Human novel trypsin inhibitor-like protein, NOV-4a encoding DNA.	
XX		
KW	Human; NOV-X protein; KIAA1233-like protein; STE20-like protein; tumour;	
KW	trypsin inhibitor-like protein; gene therapy; haematopoietic; illness;	
KW	immunological disorder; neurodegenerative disorder; Alzheimer's disease;	
KW	Parkinson's disease; immunomodulatory; pharmacogenomic; haemostatic;	
KW	human immunodeficiency virus; HIV; fertility disorder; neuroprotective;	
KW	cytostatic; nontropic; anti-fertility; cancer; NOV-4a protein; ds.	
XX		
OS	Homo sapiens.	
XX		
Key		
XX		
FT	Location/Qualifiers	
FT	453..1604	
FT	/tag=a	
FT	/product="Human novel trypsin inhibitor-like protein,	
XX	NOV-4a"	
XX		
PN	W0200162928-A2.	
XX		
PD	30-AUG-2001.	
XX		
PF	26-FEB-2001; 2001WO-US06151.	
XX		
PR	25-FEB-2000; 2000US-0184951.	
PR	28-FEB-2000; 2000US-0185348.	
PR	01-MAR-2000; 2000US-0185967.	
XX		

PR 18-APR-2000; 2000US-0197723.
PR 27-APR-2000; 2000US-0199957.
PR 23-FEB-2001; 2001US-0789390.
XX
XX
PA (CURA-) CURAGEN CORP.
XX
XX
PI Vernet CAM, Fernandes E, Shinkets RA, Macdougall J, Spaderna SK,
XX
XX WPI; 2001-582051/65.
DR P-PSDB; AAE10615.
DR
XX
XX
PT New isolated K18A1233-like, SME20-like, or trypsin inhibitor-like
PT polypeptide for diagnosing and treating pathological disorders, such as
PT Parkinson's disease and for use in pharmacogenomics -
XX
XX
PS Claim 9; Page 80-81; 189pp; English.

The invention relates to novel human polypeptides referred as NOV-X and their corresponding nucleic acid sequences. NOV-X collectively include NOV-1, NOV-2a and NOV-2b which are novel KIAA1233-like polypeptides, NOV-3a, NOV-3b, NOV-3c and NOV-3d which are novel SMP20-like polypeptides and NOV-4a, NOV-4b, NOV-4c, NOV-4d and NOV-4e which are novel trypsin inhibitor-like polypeptides. NOV-X is used to identify a potential therapeutic agent that can modulate its activity and can be used for treating a pathology related to aberrant expression or aberrant physiological interactions of NOV-X. NOV-X or its DNA is used to determine the presence or predisposition to a disease associated with altered levels of NOV-X. NOV-X, its DNA and its antibody are used to treat or prevent a pathology associated with NOV-X. The pathological states that can be treated or prevented are haematopoietic, cancer, immunological, tumour, neurodegenerative (e.g. Alzheimer's and Parkinson's diseases), human immunodeficiency virus (HIV) illness and fertility disorders. NOV-X and its DNA are used in pharmacogenomics for predictive medicine. NOV-X DNA is used in gene therapy. The present sequence is a DNA encoding human novel trypsin inhibitor-like protein, NOV-4a.

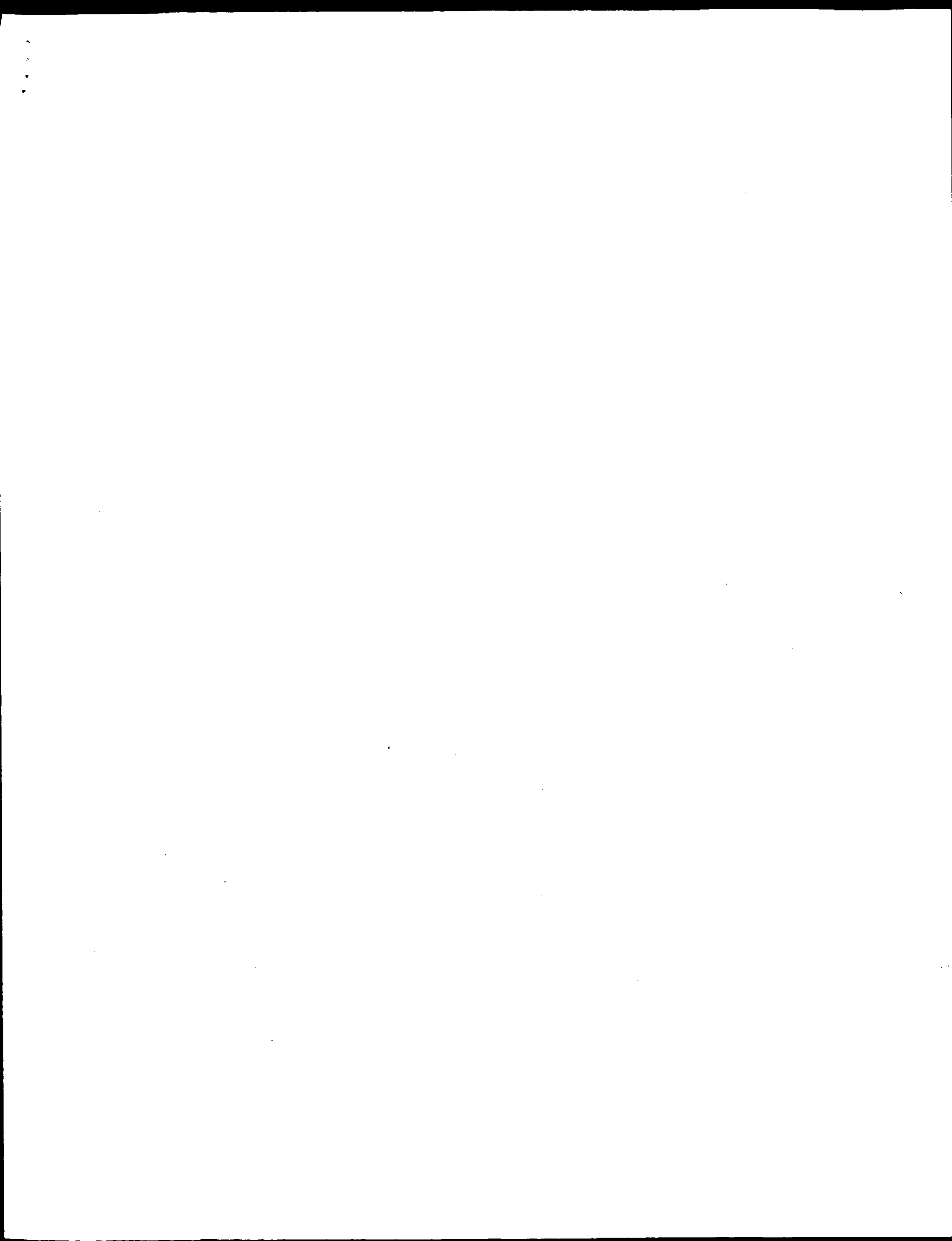
Sequence 2305 BP; 542 A; 619 C; 647 G; 497 T; 0 other;

Query Match	Similarity	Score	ID	Length
Best Local	93.4%	1286.2	DB 22	2205
Matches 1333	Conservative	0	Mismatches	3
			Indels	95
			Gaps	1

Dh	531	ccccaccgagagagctgcaaacccctcgtgctccagagaggtgcctcgggacctatgctcacg	590
Qy	481	caactaacacacagatgctctgggcccacacacaaagatcggcttgctgtgaaacacctgc	540
Dh	591	caactaacacagatgcttggtggccacacacaaagatcggcttgctgtgaaacacctgc	650
Qy	541	cggagagatgactcgtctcggggagaaatttgggagaaacggtgtactcttgctgcaattat	600
Dh	601	tcccacaaggaggaaactggaattggaagaacccctcaacaaagaaatggccctcctcgaag	660
Qy	711	tctcccaaaagggaactcgtgagatggaaagcctccatacaagaatggccctcgtctcgaag	770
Dh	661	tgcaccaaccagctatgagagagcagctgcagagaaacctggttaccgagaagaacctac	720
Qy	771	tgcaccaaccagctatgagagagcagctgcagagaaacctggttaccgagaagaacctac	830
Dh	721	accccaaaaacctgbaaacggagacgagatgagatggagagaaacggtcccatctcgaagaa	780
Qy	831	accccaaaaacctgbaaacggagacgagatgagatggagagaaacggtcccatctcgaagaa	890
Dh	781	aaacgatgtttgggtctccaaacggagggttgatgagaccacaaagcccaagaaacctctgcg	840
Qy	891	aaacgatgtttgggtctccaaacggagggttgatgagaccacaaagcccaagaaacctctgcg	950
Dh	841	gtcaactacatgaccccaagtctgtcaagatgtgacacacaagaatggaaggagacaggtgcagaagg	900
Qy	951	gtcaactacatgaccccaagtctgtcaagatgtgacacacaagaatggaaggagacaggtgcagaagg	1010
Dh	901	tccacgtgttacaacgggtacacaagctgcaccagacagagctgcctggaacacaaagcgcgaagaacctt	960
Qy	1011	tccacgtgttacaacgggtacacaagctgcaccagacagagctgcctggaacacaaagcgcgaagaacctt	1070
Dh	961	ggaagatctgtcttcatagaagctcgtcttaagcatatgcgcgcgcgcgcacatccataagggatc	1020
Qy	1071	ggaagatctgtcttcatagaagctcgtcttaagcatatgcgcgcgcgcgcacatccataagggatc	1130
Dh	1021	ctgagatcaacaaggagagagcctcgtgtgagatataccaacaaagaaagtgctccctcttcctg	1080
Qy	1131	ctgagatcaacaaggagagagcctcgtgtgagatataccaacaaagaaagtgctccctcttcctg	1190
Dh	1081	aagctctgagagacacagcgtgtgcagtccctccagcacaatacaaaccttccagctcatcatcgt	1140
Qy	1191	aagctctgagagacacagcgtgtgcagtccctccagcacaatacaaaccttccagctcatcatcgt	1250
Dh	1141	gtgtgcacaagaatggaaggtgacagagatttgagactcttaacagacacggtgtgcacagtgtgcgcg	1200
Qy	1251	gtgtgcacaagaatggaaggtgacagagatttgagactcttaacagacacggtgtgcacagtgtgcgcg	1310
Dh	1201	tttgaaaaagccagcgaactcaactctgcccagaatccattgtgtccggacacatgcaagaagcga	1260
Qy	1311	tttgaaaaagccagcgaactcaactctgcccagaatccattgtgtccggacacatgcaagaagcga	1370
Dh	1261	ccttcctactcgtggtcctcgggtgttttggaacaacaattatagcagatatccctcaagacatcgc	1320
Qy	1371	ccttcctactcgtggtcctcgggtgttttggaacaacaattatagcagatatccctcaagacatcgc	1430
Dh	1321	aagaacagctgtgcacgcggggagatcatcatcagaacaaagagagtggtggtgacgtgtgacgtgtgag	1380
Qy	1431	aagaacagctgtgcacgcggggagatcatcatcagaacaaagagagtggtggtgacgtgtgacgtgtgag	1490
Dh	1381	cccgctgataaaagaagaacctacggtggtcctgcaggaatggaggttttaagctctgaagaag	1440
Qy	1491	cccgctgataaaagaagaacctacggtggtcctgcaggaatggaggttttaagctctgaagaag	1550
Dh	1441	ctgggggaactcctcgggagatggaagaagccttcacggatcttgctgtcgaagcgaag	1491
Qy	1551	ctgggggaactcctcgggagatggaagaagccttcacggatcttgctgtcgaagcgaag	1601
RESULT 11			

XX	AAH15690	standard; cDNA; 1690 BP.
XX	AAH15690:	
XX	26-JUN-2001	(first entry)
XX	Human cDNA sequence	SEQ ID NO:14058.
XX	Human:	primer: detection; diagnosis; antisense therapy; gene therapy; ss.
XX	Homo sapiens.	
XX	EP1074617-A2.	
XX	07-FEB-2001.	
XX	28-JUL-2000;	2000EP-0116126.
XX	29-JUN-1999;	99JP-0248036.
XX	27-AUG-1999;	98JP-0300283.
XX	11-JAN-2000;	2000JP-0118776.
XX	02-MAY-2000;	2000JP-0183767.
XX	09-JUN-2000;	2000JP-0241899.
XX	(HELI-) HELIX RES INST.	
XX	Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;	
XX	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;	
XX	WPI: 2001-318749/34.	
XX	Primer sets for synthesizing polynucleotides, particularly the 5602	
XX	full-length cDNAs defined in the specification, and for the detection	
XX	and/or diagnosis of the abnormality of the proteins encoded by the	
XX	full-length cDNAs -	
XX	Claim 8: SEQ ID 14058; 2537Pp + CD ROM; English.	
XX	The present invention describes primer sets for synthesizing 5602	
XX	full-length cDNAs defined in the specification. Where a primer set	
XX	comprises: (a) an oligo-dT primer and an oligonucleotide complementary	
XX	to the complementary strand of a polynucleotide which comprises one of	
XX	the 5602 nucleotide sequences defined in the specification, where the	
XX	oligonucleotide comprises at least 15 nucleotides; or (b) a combination	
XX	of an oligonucleotide comprising a sequence complementary to the -	
XX	complementary strand of a polynucleotide which comprises a 5'-end	
XX	sequence and an oligonucleotide comprising a sequence complementary to a	
XX	polynucleotide which comprises a 3'-end sequence, where the	
XX	oligonucleotide comprises at least 15 nucleotides and the combination of	
XX	the 5'-end sequence/3'-end sequence is selected from those defined in	
XX	the specification. The primer sets can be used in antisense therapy and	
XX	in gene therapy. The primers are useful for synthesizing polynucleotides,	
XX	particularly full-length cDNAs. The primers are also useful for the	
XX	detection and/or diagnosis of the abnormality of the proteins encoded by	
XX	the full-length cDNAs. The primers allow obtaining of the full-length	
XX	cDNAs easily without any specialised methods. AAH03166 to AAH13628 and	
XX	AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to	
XX	AAB93893 represent human amino acid sequences; and AAH13629 to AAH13632	
XX	represent oligonucleotides, all of which are used in the exemplification	
XX	of the present invention.	
XX	Sequence 1690 BP; 410 A; 469 C; 489 G; 322 T; 0 other;	
XX	Query Match	74.4%; Score 1109.6; DB 22; Length 1690;
XX	Best Local Similarity: 99.6%;	Pred. No. 5, 9e-293;
XX	Matches 1112; Conservative	0; Mismatches 4; Indels 0; Gaps 0.
XX	1 atagactgcgtctgggtgggtgtcaccaccttggggctcgtctgttcgttcagcagatcc 60	
XX		
XX	192 atagactgcgtctctgggtgtcaccaccttggggctcgtctgttccttcgttcgcgagatcc 251	

OY 241 acctgggttgagcaactggaagaatctgcctgcagcgtgaggccagctcaagtctacatctggag 300
 Db 456 acctgggttgagcaactggaagaatctgcctgcagcgtgaggccagctcaagtctacatctggag 515
 OY 301 caggggccaccagctctgctggtgtgtccatcgggcagaacctgggcgcctcaactggggcaga 360
 Db 516 caggggccaccagctctgctggtgtgtccatcgggcagaacctgggcgcctcaactggggcaga 575
 OY 361 tatgcctctcgggggttccatgtgcagctcctgtatgacgaagtgaaagatactacactac 420
 Db 576 tatgcctctcgggggttccatgtgcagctcctgtatgacgaagtgaaagatactacactac 635
 OY 421 cccctaccggagcgggtgtgcaacccctgtgtgtccagagagagtgctccgggacctatgtgcag 480
 Db 636 cccctaccggagcgggtgtgcaacccctgtgtgtccagagagagtgctccgggacctatgtgcag 695
 OY 481 cactacacacagatagtttgggccacacacacaagaatcggtgtgtgtctgtgaacctgtc 540
 Db 696 cactacacacagatagtttgggccacacacacaagaatcggtgtgtgtctgtgaacctgtc 755
 OY 541 cggaaagatgactgtctggaggaagatcttggagagaacgcgcttactcttctgcgaatat 600
 Db 756 cggaaagatgactgtctggaggaagatcttggagagaacgcgcttactcttctgcgaatat 815
 OY 601 tctccaaagggaactgtgattggagaagccccctacaagaatgaccggccctctgtga 660
 Db 816 tctccaaagggaactgtgattggagaagccccctacaagaatgaccggccctctgtga 875
 OY 661 tggccaccacacatctggaggcagctgtgcagaaacaacttgttaccgaaagaactac 720
 Db 876 tggccaccacacatctggaggcagctgtgcagaaacaacttgttaccgaaagaactac 935
 OY 721 actc 724
 Db 936 actc 939
 RESULT 13
 AAH05058
 ID AAH05058 standard; cDNA; 840 BP.
 XX
 AC AAH05058;
 XX
 DT 26-JUN-2001 (first entry)
 DE
 XX Human cDNA clone (5'-primer) SEQ ID NO:1893.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss
 OS Homo sapiens.
 XX
 KW
 XX
 PN EP1074617-A2.
 PD
 PD 07-FEB-2001.
 PE 28-JUL-2000; 2000EP-0116126.
 XX
 PR 29-JUL-1999; 99DP-0248036.
 PR 27-AUG-1999; 99DP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Oka T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI: 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the




```

Db 61 MLHNKLRGVOPOASNMWYMTWDELKSAAMASOCIWEHGPSTSLVSTIGONLGAHMR 120
QY 121 YRSPGFHVSMTDEVKDITYPPECECPNCPCEPSCGPMCTHTYTOIWMATTNKIGCAVNTC 180
Db 121 YRSPGFHVSMTDEVKDITYPPECECPNCPCEPSCGPMCTHTYTOIWMATTNKIGCAVNTC 180
QY 181 RKMTWGEWENAVYFVNCNPKGNWIGEPYKNGRCPSCPPSYGSGCRNNLCYREBTY 240
Db 181 RKMTWGEWENAVYFVNCNPKGNWIGEPYKNGRCPSCPPSYGSGCRNNLCYREBTY 240
QY 241 TPKEPTDENNEVEETAPRPEENHVMLOPRVMPRTPKKTSANVMTOVRCDTKMKDRCKG 300
Db 241 TPKEPTDENNEVEETAPRPEENHVMLOPRVMPRTPKKTSANVMTOVRCDTKMKDRCKG 300
QY 301 STCNRYOCPAGCLNHNKAKIFGLTFYESSSSICRAAIHYGLDDKGLVDITRNGKVPFV 360
Db 301 STCNRYOCPAGCLNHNKAKIFGLTFYESSSSICRAAIHYGLDDKGLVDITRNGKVPFV 360
QY 361 KSERHGVOSLSKYKSSPSSFWKVKVODLCTTYAOUCPEKPAHCPRIHCPAHCKDE 420
Db 361 KSERHGVOSLSKYKSSPSSFWKVKVODLCTTYAOUCPEKPAHCPRIHCPAHCKDE 420
QY 421 PSYMAVPEGNITADTSSICKTAVHAGVISNESGDDVMPVDDKKTYSLSRNGVOS 480
Db 421 PSYMAVPEGNITADTSSICKTAVHAGVISNESGDDVMPVDDKKTYSLSRNGVOS 480
QY 481 LGTPRDGKAFRIFAVRQ 497
Db 481 LGTPRDGKAFRIFAVRQ 497

```

```

RESULT 2
ID 096K61 PRELIMINARY; PRT: 371 AA.
AC 096K61;

```

```

DI 01-DEC-2001 (Tremblrel. 19, Created)
DI 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DI 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE CDNA FLJ14469 FIS CLONE MAMMA1002881, WEAKLY SIMILAR TO GLIOMA
DE PATHOGENESIS-RELATED PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN NCBI_TaxID=9606;
RP SEQUENCE FROM N.A.
RC TISSUE=MAMMARY GLAND.
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuho Y., Sasaki N.;
RT "NED0 human cDNA sequencing project."
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK027395; BAB5081.1;
SQ SEQUENCE 371 AA; 42207 MW; 8D23FBEL4F53E85C CRC64;

```

```

Query Match
Best Local Similarity 75.1%; Score 2071; DB 4; Length 371;
Matches 369; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 MSCVIGGVIPGLGLFVCGSGVLLPNTYLLLELLSKYOHNSHRYRAIPREDKEITL 60
Db 1 MSCVIGGVIPGLGLFVCGSGVLLPNTYLLLELLSKYOHNSHRYRAIPREDKEITL 60
QY 61 MLHNKLRGVOPOASNMWYMTWDELKSAAMASOCIWEHGPSTSLVSTIGONLGAHMR 120
Db 61 MLHNKLRGVOPOASNMWYMTWDELKSAAMASOCIWEHGPSTSLVSTIGONLGAHMR 120
QY 121 YRSPGFHVSMTDEVKDITYPPECECPNCPCEPSCGPMCTHTYTOIWMATTNKIGCAVNTC 180
Db 121 YRSPGFHVSMTDEVKDITYPPECECPNCPCEPSCGPMCTHTYTOIWMATTNKIGCAVNTC 180
QY 181 RKMTWGEWENAVYFVNCNPKGNWIGEPYKNGRCPSCPPSYGSGCRNNLCYREBTY 240
Db 181 RKMTWGEWENAVYFVNCNPKGNWIGEPYKNGRCPSCPPSYGSGCRNNLCYREBTY 240
QY 241 TPKEPTDENNEVEETAPRPEENHVMLOPRVMPRTPKKTSANVMTOVRCDTKMKDRCKG 300
Db 241 TPKEPTDENNEVEETAPRPEENHVMLOPRVMPRTPKKTSANVMTOVRCDTKMKDRCKG 300
QY 301 STCNRYOCPAGCLNHNKAKIFGLTFYESSSSICRAAIHYGLDDKGLVDITRNGKVPFV 360
Db 301 STCNRYOCPAGCLNHNKAKIFGLTFYESSSSICRAAIHYGLDDKGLVDITRNGKVPFV 360
QY 361 KSERHGVOSLSKYKSSPSSFWKVKVODLCTTYAOUCPEKPAHCPRIHCPAHCKDE 420
Db 361 KSERHGVOSLSKYKSSPSSFWKVKVODLCTTYAOUCPEKPAHCPRIHCPAHCKDE 420
QY 421 PSYMAVPEGNITADTSSICKTAVHAGVISNESGDDVMPVDDKKTYSLSRNGVOS 480
Db 421 PSYMAVPEGNITADTSSICKTAVHAGVISNESGDDVMPVDDKKTYSLSRNGVOS 480
QY 481 LGTPRDGKAFRIFAVRQ 497
Db 481 LGTPRDGKAFRIFAVRQ 497

```

```

QY 181 RKMTWGEWENAVYFVNCNPKGNWIGEPYKNGRCPSCPPSYGSGCRNNLCYREBTY 240
Db 181 RKMTWGEWENAVYFVNCNPKGNWIGEPYKNGRCPSCPPSYGSGCRNNLCYREBTY 240
QY 241 TPKEPTDENNEVEETAPRPEENHVMLOPRVMPRTPKKTSANVMTOVRCDTKMKDRCKG 300
Db 241 TPKEPTDENNEVEETAPRPEENHVMLOPRVMPRTPKKTSANVMTOVRCDTKMKDRCKG 300
QY 301 STCNRYOCPAGCLNHNKAKIFGLTFYESSSSICRAAIHYGLDDKGLVDITRNGKVPFV 360
Db 301 STCNRYOCPAGCLNHNKAKIFGLTFYESSSSICRAAIHYGLDDKGLVDITRNGKVPFV 360
QY 361 KSERHGVOSLSKYKSSPSSFWKVKVODLCTTYAOUCPEKPAHCPRIHCPAHCKDE 420
Db 361 KSERHGVOSLSKYKSSPSSFWKVKVODLCTTYAOUCPEKPAHCPRIHCPAHCKDE 420

```

```

RESULT 3
ID 09D2R3 PRELIMINARY; PRT: 434 AA.
AC 09D2R3;
DI 01-JUN-2001 (Tremblrel. 17, Created)
DI 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DI 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE 1810049K24RIK PROTEIN.
GN 1810049K24RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN NCBI_TaxID=10090;
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=PANCREAS.
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamada S.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fiedischmann W., Gaasterland T., Giesl C., King B., Kochwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsi G.,
RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Holman M., Hume D.A., Kamiya M., Lee N.H.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitaker C., Wilming L.,
RA Wyshak-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohlsauki S.,
RT "functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
DR EMBL; AK010034; BAB31519.1;
DR HSSP; P04284; ICFF.
DR MGD; MGI:1926142; 1810049K24RIK.
DR InterPro; IPR004043; ICCL.
DR Pfam; PF00188; SCP.1.
DR PRINTS; PR00837; V5PXLKE.
DR SMART; SM00198; SCP.1.
DR PROSITE; PS01010; SCP_AGS_PRL_SCT_2; 1.
SQ SEQUENCE 434 AA; 48593 MW; C25067E814647AC0 CRC64;

```

```

Query Match
Best Local Similarity 72.0%; Score 1987.5; DB 11; Length 434;
Matches 348; Conservative 37; Mismatches 49; Indels 3; Gaps 1;

```

```

QY 61 MLHNKLRGVOPOASNMWYMTWDELKSAAMASOCIWEHGPSTSLVSTIGONLGAHMR 120
Db 61 MLHNKLRGVOPOASNMWYMTWDELKSAAMASOCIWEHGPSTSLVSTIGONLGAHMR 120
QY 121 YRSPGFHVSMTDEVKDITYPPECECPNCPCEPSCGPMCTHTYTOIWMATTNKIGCAVNTC 180
Db 121 YRSPGFHVSMTDEVKDITYPPECECPNCPCEPSCGPMCTHTYTOIWMATTNKIGCAVNTC 180
QY 181 RKMTWGEWENAVYFVNCNPKGNWIGEPYKNGRCPSCPPSYGSGCRNNLCYREBTY 240
Db 181 RKMTWGEWENAVYFVNCNPKGNWIGEPYKNGRCPSCPPSYGSGCRNNLCYREBTY 240
QY 241 TPKEPTDENNEVEETAPRPEENHVMLOPRVMPRTPKKTSANVMTOVRCDTKMKDRCKG 300
Db 241 TPKEPTDENNEVEETAPRPEENHVMLOPRVMPRTPKKTSANVMTOVRCDTKMKDRCKG 300
QY 301 STCNRYOCPAGCLNHNKAKIFGLTFYESSSSICRAAIHYGLDDKGLVDITRNGKVPFV 360
Db 301 STCNRYOCPAGCLNHNKAKIFGLTFYESSSSICRAAIHYGLDDKGLVDITRNGKVPFV 360
QY 361 KSERHGVOSLSKYKSSPSSFWKVKVODLCTTYAOUCPEKPAHCPRIHCPAHCKDE 420
Db 361 KSERHGVOSLSKYKSSPSSFWKVKVODLCTTYAOUCPEKPAHCPRIHCPAHCKDE 420
QY 421 PSYMAVPEGNITADTSSICKTAVHAGVISNESGDDVMPVDDKKTYSLSRNGVOS 480
Db 421 PSYMAVPEGNITADTSSICKTAVHAGVISNESGDDVMPVDDKKTYSLSRNGVOS 480
QY 481 LGTPRDGKAFRIFAVRQ 497
Db 481 LGTPRDGKAFRIFAVRQ 497

```

DB 1 MHNKRGVYTPPASNMEHTWDELESAANAHAHRCLEMEHGPAGLLRSIGONLAVHGR 60
 QY 121 YSPGPHVQSMYDEVDKDYTPYPPSECNPCWPCBRCGPMCTHTYQIWAHTTKICAVNTC 180
 DB 61 YRSPGPHVQSMYDEVDKDYTPYPPSECNPCWPCBRCGPMCTHTYQIWAHTTKICAVNTC 120
 QY 181 RKATVGEVMEANAYTVVCYSPKGNMIGEPAPYKNGRPSCEPPSYGSGCRNNLCYRETY 240
 DB 121 RNNVWGDVMEANAYTVVCYSPKGNMIGEPAPYKNGRPSCEPPSYGSGCRNNLCYRETY 180
 QY 241 TPKPETDEMENEYTAIPPEENHVLQPRVMPRTKPKTSANVTQVYKCDTKMKDRCKG 300
 DB 181 KHKPEVDMMNEVESPPAPRETHWQPRVY---KTKETPEVINFMTQVYHCDTKMKDRCKG 237
 QY 301 STCNROCPAGCLNKAHAKIFGSLFYESSSSICRAAIHYGLIDDKGLVDITRNGKVPFV 360
 DB 238 STCNROCPAGCLNKAHAKIFGSLFYESSSSICRAAIHYGLIDDKGLVDITRNGKVPFV 297
 QY 361 KSERHGVQSLSKYKPSSEFWYKVKVODDCTYTVVQALCPFEKATHTCPRIHCPAHKDE 420
 DB 298 KSKNGMESLSKYPSSFTVSKVETETAVDCHATVQALCPFEKATHTCPRIHCPAHKDE 357
 QY 421 PSYAPVPGINITYADTSSICKTAHAGVYISNESGDDVMPVDPKKTYYGSLRNGVQSES 480
 DB 358 PSYAPVPGINITYADTSSICKTAHAGVYISNESGDDVMPVDPKKTYYGSLRNGVQSES 417
 QY 481 LGTPPDGKAFRIFAVRQ 497
 DB 418 LNTQNGNARIFRAVRQ 434

RESULT 4
 09H336 PRELIMINARY; PRT; 500 AA.
 ID 09H336:
 AC 01-MAR-2001 (TREMblrel. 16, Created)
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE PUTATIVE SECRETORY PROTEIN PRECURSOR (COCOACRISP).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 NC NCBL_TaxID-9606;
 RX NCBL_TaxID-9606;
 RN [1]
 RC TISSUE-MORTA;
 RA Zhao Y., Cao H., Jiang Y., Meng X., Zhao X., Liu D., Ding J.;
 RA "Cloning and characterization of a trypsin inhibitor-like protein gene
 in human aorta."
 RT Submitted (Apr-1999) to the EMBL/GenBank/DBJ databases.
 RL [2]
 RN SEQUENCE FROM N.A.
 RP TISSUE-HEART, AND TESTIS;
 RA Smith D.M., Collins-Racie L.A., Lavallee E.R., Gerner L., Roberts D.J.,
 RA Mario V.A., Copeland N.G., Jenkins N.A., McCoy J., Tabin C.J.;
 RA "A novel cysteine-rich secreted protein (CRISP) family member,
 RT Cococristp, provides insight into the process of septation in the
 RT developing chicken midbrain."
 RT Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 RL EMBL: AF142573; AAC3287.1;
 DR EMBL: AF142573; AAC3287.1;
 DR HSSP: P04284; ICFE.
 DR InterPro: IPR004043; LCCL.
 DR InterPro: IPR001283; SCP.
 DR Pfam: PF00188; SCP.1.
 DR PRINTS: PR00837; V5TPXLIKE.
 DR SMART: SM00198; SCP.1.
 DR PROSITE: PS01010; SCP_AGS_PRL_SCT_2; 1.
 DR PROSITE: PS01010; SCP_AGS_PRL_SCT_2; 1.
 SEQUENCE 500 AA; 56888 MW; 203BIDC4DDA003CB CRC64;

Query Match 59.5%; Score 1640.5; DB 4; Length 500;
 Best Local Similarity 59.5%; Pred. No. 9.5e-151;

Matches 292; Conservative 74; Mismatches 112; Indels 13; Gaps 7;
 QY 13 LFLVCGSGGILLPVNTLLELLSTYQNF-----SHSRVRAIPREDKEELMLHNRK 67
 DB 14 VLFMARAIPIAMVVPVATLLEKLEKEDMDGEMWIAKORGRRAITDNDMSGILDLHNRK 73
 QY 68 GQVQPOASNMETWMDDELEKSAAMASOCIWHEGPTSLVYSGONIGAHGRTYRSPGFH 127
 DB 74 SQVPTASNMETWMDDELEKSAAMASOCIWHEGPTSLVYSGONIGAHGRTYRSPGFH 133
 QY 128 VQSMYDEVDKDYTPYPPSECNPCWPCBRCGPMCTHTYQIWAHTTKICAVNTC 187
 DB 134 VQSMYDEVDKDYTPYPPSECNPCWPCBRCGPMCTHTYQIWAHTTKICAVNTC 193
 QY 188 EWMENAYFYVCNYSKPKGNMIGEPAPYKNGRPSCEPPSYGSGCRNNLCYRETY 245
 DB 194 QIMPKAVYLVCCNYSKPKGNMIGEPAPYKNGRPSCEPPSYGSGCRNNLCYRETY 252
 QY 246 TDEMENEYTAIP-EEENHVLQPRVMPRTKPKTSANVTQVYKCDTKMKDRCKG 304
 DB 253 EEFTELEEROOSQVMDTHV--RTRSDSSRNEVISA-QQMSQIVSEVLRDQCKGTTCN 309
 QY 305 RYOCPCAGCLNKAHAKIFGSLFYESSSSICRAAIHYGLIDDKGLVDITRNGKVPFV 364
 DB 310 RYECPCAGCLNKAHAKIFGSLFYESSSSICRAAIHYGLIDDKGLVDITRNGKVPFV 369
 QY 365 HGVQSLSKYKPSSEFWYKVKVODDCTYTVVQALCPFEKATHTCPRIHCPAHKDE 424
 DB 370 NGIOTIGKQVANSFTVSKVETETAVDCHATVQALCPFEKATHTCPRIHCPAHKDE 429
 QY 425 APVFGTINITYADTSSICKTAHAGVYISNESGDDVMPVDPKKTYYGSLRNGVQSES 484
 DB 430 ARVIGTRVYSDLSICRAAHAGVYISNESGDDVMPVDPKKTYYGSLRNGVQSES 488
 QY 485 RDGKAFRIFAVR 495
 DB 489 PGKAFRIFAVR 499

RESULT 5
 09H336 PRELIMINARY; PRT; 523 AA.
 ID 09H336:
 AC 01-JUN-2001 (TREMblrel. 17, Created)
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE COCOACRISP.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 NC NCBL_TaxID-9031;
 RX NCBL_TaxID-9031;
 RN [1]
 RC SEQUENCE FROM N.A.
 RA Smith D.M., Collins-Racie L.A., Lavallee E.R., Gerner L., Roberts D.J.,
 RA Mario V.A., Copeland N.G., Jenkins N.A., McCoy J., Tabin C.J.;
 RA "A novel cysteine-rich secreted protein (CRISP) family member,
 RT Cococristp, provides insight into the process of septation in the
 RT developing chicken midbrain."
 RT Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 RL EMBL: AF329199; AAK16497.1;
 DR EMBL: AF329199; AAK16497.1;
 DR HSSP: P04284; ICFE.
 DR InterPro: IPR004043; LCCL.
 DR InterPro: IPR001283; SCP.
 DR Pfam: PF00188; SCP.1.
 DR PRINTS: PR00837; V5TPXLIKE.
 DR SMART: SM00198; SCP.1.
 DR PROSITE: PS01010; SCP_AGS_PRL_SCT_2; 1.
 DR PROSITE: PS01010; SCP_AGS_PRL_SCT_2; 1.
 SEQUENCE 523 AA; 59049 MW; 3AF283C19DC3C55F CRC64;

Query Match 58.5%; Score 1613; DB 13; Length 523;
 Best Local Similarity 58.2%; Pred. No. 4.8e-148;

[illegible]

RESULT	7		
Q9Z006		PRELIMINARY;	PRT; 188 AA.
ID	Q9Z006		
DT	01-MAY-1999 (TREMBLrel. 10, Created)		
DT	01-MAY-1999 (TREMBLrel. 10, Last sequence update)		
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)		
DE	LATE GESTATION LUNG PROTEIN 1,		
GN	LG1,		
GN	Rattus norvegicus (Rat),		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus		
ON	NCHI_TaxID=10116;		
ON	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=WISTAR; TISSUE=LUNG;		
RX	MEDLINE=99292450; PubMed=10362728;		
RA	Kaplan F., Ledoux P., Kassamali F.O., Gagnon S., Post M., Koehler D.,		
RA	Deinling J., Sweezey N.B.;		
RA	"A novel developmentally regulated gene in lung mesenchyme: homology		
RL	to a tumor derived trypsin inhibitor ";		
RL	Am. J. Physiol. 276:L1027-L1036(1999).		
DR	EMBL; AF109674; AAD16986.1; .		
DR	HSSP; P04284; ICFE.		
DR	InterPro: IPR001283; SCP.		
DR	Pfam: PF00188; SCP; 1.		
DR	PRINTS; PR00837; V5TPYLKE.		
DR	SMART; SM00198; SCP; 1.		
DR	PROSITE; PS01010; SCP_AGS_PRL_SC7_2; 1		
DR	SEQUENCE 188 AA: 21907 MW; 586CC77515360BCE CRC64;		

Query Match	55.7%	Score 1536.5	DB 11	Length 507
Best Local Similarity	55.4%	Prod. No. 1.2e-140		
Matches 274; Conservative	74	Mismatches 112	Indels 35	Gaps 7

Query Match	35.1%	Score 968;	DB 11;	Length 188;
Best Local Similarity	86.5%	Prod. No. 5.9e-86;		
Matches 160;	Conservative 10;	Mismatches 15;	Indels 0;	Gaps 0

RA Kanamori T., Miyazaki K.;
 RT "cDNA cloning of a novel trypsin inhibitor with similarity to
 RT pathogenesis-related proteins, and its frequent expression in human
 RT brain cancer cells.";
 RL Blochim. Biophys. Acta 1395:202-208(1998).
 DR EMBL; D45027; BAA25066.1; -.
 DR HSSP; P04284; ICFE.
 DR InterPro: IPR001283; SCP.
 DR Pfam: PF00188; SCP; 1.
 DR PRINTS; PR00837; V5TPXLIKE.
 DR SMART; SM00198; SCP; 1.
 DR PROSITE; PS01010; SCP_AGS_PRL_SCT-2; 1.
 SQ SEQUENCE 258 AA; 29065 MW; 1915A5831637795F CRC64;

Query Match 26.8%; Score 739.5; DB 4; Length 258;
 Best Local Similarity 56.4%; Pred. No. 1.5e-63;
 Matches 133; Conservative 30; Mismatches 60; Indels 13; Gaps 2;

QY 13 LIF-LVCGSGGYLLPNTLLEELLSKYOHNES-----HSRYRAIPREDKEEI 59
 DB 11 LIFSLICEASTYVLLNSTDSSPTNFTDIEALKAQLDSADIPAKRRRTISQNDMAI 70
 QY 60 LMLHKLKRGQVOPQASNMEXYTWDELEKSAAMASQCIWEHPTSLVSTIGONGAHWG 119
 DB 71 LDYHNGVRGKVPFPAANMEYMWDEILAKSADAMATCTMDHGPSYLLRFLGONLSVFTG 130
 QY 120 KRSPGFHVSQWYDEKQDTYPRSCNWCPCRGSCPGCTHYTOIVATNTKKICAVNT 179
 DB 131 RYRSITQIVKPMYDEKDAFPYPODCNPRCPGCPGCTHYTOIVATNTSRICCAHT 190
 QY 180 CRMTWGEVWENAVFYVCNYSPPKNGWIGEAIFYKNGRCPSCPPSYGSGCRNNLCY 235
 DB 191 CQMNWVSGVMWRAVYVLCNVAPKGMWIGEAIFYKNGRCPSCPPSYGSGCTNDLCP 246

RESULT 11
 Q9H3Y0 PRELIMINARY; PRT; 253 AA.
 AC Q9H3Y0;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE DJ881L22.3 (NOVEL PROTEIN SIMILAR TO A TRYPSIN INHIBITOR).
 GN DJ881L22.3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lloyd D.;
 RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL117382; CAC04190.1; -.
 DR HSSP; P04284; ICFE.
 DR InterPro: IPR001283; SCP.
 DR Pfam; PF00188; SCP; 1.
 DR PRINTS; PR00837; V5TPXLIKE.
 DR SMART; SM00198; SCP; 1.
 SQ SEQUENCE 253 AA; 28605 MW; 28976C6F32E3887E CRC64;

Query Match 23.4%; Score 646.5; DB 4; Length 253;
 Best Local Similarity 47.7%; Pred. No. 1.6e-54;
 Matches 114; Conservative 38; Mismatches 66; Indels 21; Gaps 4;
 QY 12 GLLEFLVCGS-QGYLLPNT-----LLEEL-LSKYOHNESHSVRRAIPREDKE 57
 DB 12 GLLEFMAGQAVNALIMPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 64
 QY 58 EILMLHKLKRGQVOPQASNMEXYTWDELEKSAAMASQCIWEHPTSLVSTIGONGAH 117
 DB 65 ALLDYHNHTRASYPPPAANMEYMWDRRLAARAAEWATQCIWAHGPSQIMRYVGNLSIH 124

QY 118 MGRYSPGFHVSQWYDEKQDTYPRSCNWCPCRGSCPGCTHYTOIVATNTKKICAV 177
 DB 125 SGQYRSYVDLMKSKSEKMHYLPAPDCNPHCFWRCGDGFTCSHYDMWASSNRIGCAI 184
 QY 178 NTCRMTWGEVWENAVFYVCNYSPPKNGWIGEAIFYKNGRCPSCPPSYGSGCRNNLCY 236
 DB 185 HTCSISVWGNTWRAVYVLCNVAPKGMWIGEAIFYKNGRCPSCPPSYGSGCRNNLCY 243

RESULT 12
 Q961B1 PRELIMINARY; PRT; 120 AA.
 AC Q961B1;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-MELANOMA.
 RA Strausberg R.;
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC007689; AA07689.1; -.
 KW Hypothetical protein.
 SQ SEQUENCE 120 AA; 13497 MW; D0B40A3149B1099C CRC64;

Query Match 23.2%; Score 639; DB 4; Length 120;
 Best Local Similarity 100.0%; Pred. No. 3.1e-54;
 Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MCVLGGVYIPGLLEFLVCGSGGYLLPNTLLEELLSKYOHNESHSVRRAIPREDKEEI 60
 DB 1 MCVLGGVYIPGLLEFLVCGSGGYLLPNTLLEELLSKYOHNESHSVRRAIPREDKEEI 60
 QY 61 MHLHKLKRGQVOPQASNMEXYTWDELEKSAAMASQCIWEHPTSLVSTIGONGAHWG 120
 DB 61 MHLHKLKRGQVOPQASNMEXYTWDELEKSAAMASQCIWEHPTSLVSTIGONGAHWG 120
 RESULT 13
 O44228 PRELIMINARY; PRT; 415 AA.
 AC O44228;
 DT 01-JUN-1998 (TREMBLrel. 06, Created)
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
 DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE HRTT-1 PROTEIN.
 GN HRTT-1.
 OS Halocynthia roretzi (Sea squirt).
 OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
 OC Stolidobranchia; Pyridae; Halocynthia.
 OX NCBI_TaxID=7729;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-98267369; PubMed-9601991;
 RA Hotta K., Takahashi H., Satoh N.;
 RT "Expression of an ascidian gene in the tip of the tail of tail-bud-
 RT stage embryos.";
 RL Dev. Genes Evol. 208:164-167(1998).
 DR EMBL; AB009609; BAA24011.1; -.
 DR HSSP; P04284; ICFE.
 DR InterPro: IPR001283; SCP.
 DR InterPro: IPR000884; TSP1.
 DR Pfam; PF00188; SCP; 1.
 DR Pfam; PF00090; TSP; 1.
 DR PRINTS; PR00837; V5TPXLIKE.
 DR SMART; SM00198; SCP; 1.

DR SMART: SM00209: TSP1, 1
DR PROSITE: PS01009: SCP_A65_PRL_SC7_1; 1.
DR PROSITE: PS01010: SCP_A65_PRL_SC7_2; 1.
DR PROSITE: PS50092: TSP1; 1.
SQ SEQUENCE 415 AA; 47365 MW; 22366B3F19DEDD3 CRC64;

Query Match
Best Local Similarity 12.9%; Score 357; DB 5; Length 415;
Matches 110; Conservative 58; Mismatches 128; Indels 96; Gaps 18;

QY 24 LIPNTLLEELLSKY-----QHNSHSRVRA-----IPREKEEL-----LM 61
DB 74 ILSEKVEEITARDIGLDPANETFS-VKAAVAKIIEGKIMDKIVLKPFEMALR 132
QY 62 LHNKRGVOVPOASMEYMTWDELKSAAMASQICWEHPT---SLVSGONLGAHW 118
DB 133 THNARRAAOPKAAKRRMTWDELELAVASRKCIEHNPRKHSFEYVGENLIST 192
QY 119 GRYRSPGF---HYOSVDEVDYTYPRSECNPCPEKCS-GPMCTHYQIWMATTKNG 174
DB 193 GYALFPLSLMKHVAEAMDEKQYDE-----TKKQKQKMGCHYTVWADTFEMG 243
QY 175 CAVNTCRKMTYGEVWENAVFYCNYSFGNMIGAPYKNGRPSCEPSSGSGCRNNLC 234
DB 244 CGYTRCSIDYRGRMRKALLVCNYGGGNY-PTHFPTVAPSCSKAPT--DICRNLG 300
QY 235 YREETYTPKPTDEMEY--ETADIPERNHYWLOPRVWRPTKPKTSAVNYMTQVRCDT 292
DB 301 -----NNVIRDLRLDRDKIKSEMTW--SSCSKSGVSGSTRERNCNT 343
QY 293 KMKDRCKG-----STCNRYOCPAGCLNHNKAKIRGSLFYESSSICRAIHYIIDKGL 347
DB 344 FVPDDCKDPSEVFCYKRC-----KAAMFGN-----GGSFYNIWMNOG-- 384
QY 348 VDIRNGKVPFVSKSRHGVOSISKYRPSSE 379
DB 385 -DKLLKSLQ-----QALQKHLGFSF 405

RESULT 14
Q969K2 PRELIMINARY; PRT; 266 AA.
AC Q969K2; 01-DEC-2001 (Tremblrel, 19, Created)
DT 01-DEC-2001 (Tremblrel, 19, last sequence update)
DE GLIOMA PATHOGENESIS-RELATED PROTEIN (UNKNOWN) (PROTEIN FOR RTVP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Wu J., Zhang B., Zhou Y., Peng X., Yuan J., Qiang B.;
RN Submitted (JUL-2001) to the EMBL/Genbank/DBJ databases.
RC TISSUE=RENAL ADENOCARCINOMA;
RA Strausberg R.;
RL Submitted (AUG-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF400440; AAK92489.1;
DR EMBL; BC012510; AAH12510.1;
SQ SEQUENCE 266 AA; 30366 MW; C0466C085EFB2B7A CRC64;

Query Match
Best Local Similarity 12.9%; Score 355; DB 4; Length 266;
Matches 82; Conservative 33; Mismatches 75; Indels 40; Gaps 12;
QY 25 LIPNTLLEELLSKYQHNSHSRVRAIPRED-KEIILNHNKRGVOVPOASMEYMTW 83

DB 5 LATIAMWSEVSNYSHT---ANILPDIEENEDFIDCVRIHNKFRSEYKPTASDKLYMTWD 61
QY 84 DELEKSAAMASQICWEHPT---PTSL---LVSGONLGAHWGRYRSPGHVQS---W 131
DB 62 PALAQIAKAMASNGQFSGHNTPLKPPHKLHFNFTSLGENT--W-TGSVPLEFVSATNW 117
QY 132 YDEKVDYTYPRSCNFWPCPEKSGPMCTHYQIWMATTKNGICAVNTCRKMTYGEVWE 191
DB 118 YDEIYDQYDF-----KTRICKRVCGHYTVWADSYKVGCAVOFCPKYSGF-DALS 166
QY 192 NAVFYCNYSFGN---WIGAPYKNGRPSCEPSSGSGCRNNLCYREE 238
DB 167 NGAHFICNYGPGGNYPTW---PYKRGATCSACPNN--DKCLDNLGVNRQ 210

RESULT 15
Q9J56 PRELIMINARY; PRT; 489 AA.
AC Q9J56; 01-OCT-2000 (Tremblrel, 15, Created)
DT 01-OCT-2000 (Tremblrel, 15, last sequence update)
DE CYSTEINE-RICH PROTEASE INHIBITOR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Jang J.S., Hahn J.H.;
RT Identification of novel mouse cysteine-rich protease inhibitor gene."
RN Submitted (JUL-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AB046537; BAB03398.1;
DR HSSP; P04284; ICEF.
DR InterPro; IPR001283; SCP.
DR Pfam; PF00188; SCP; 1.
DR PRINTS; PR00837; V5TPXLIKE.
DR SMART; SM00198; SCP; 1.
DR PROSITE; PS01009; SCP_A65_PRL_SC7_1; 1.
DR PROSITE; PS01010; SCP_A65_PRL_SC7_2; 1.
KW Protease.
SQ SEQUENCE 489 AA; 52676 MW; 01C207BEB1E3CB9D CRC64;

Query Match
Best Local Similarity 12.6%; Score 349; DB 11; Length 489;
Matches 77; Conservative 28; Mismatches 86; Indels 16; Gaps 6;
QY 50 AIPRDKKEIILNHNKRGVOVPOASMEYMTWDELKSAAMASQICWEHPTSLV 109
DB 20 ALTEDEKOTWDLHNOYTAOVSPPASDMLQKRWDELAAFAKAAOKVWGNHER--GR 77
QY 110 IGONLGAHWGRYRSPGFVOSVWYDEKVDYTYPRSECNPCPEKSGPMCTHYQIWMAT 169
DB 78 ROENLFAITDEGMVPLAVGAWMHEHEYNFS-TATCDP-----NOMCHYTVWWSK 129
QY 170 TNKIGCAVNTCRKMTYGEVWENAVFYCNYSFGNMIGAPYKNGRPSCEPSSGSGC 229
DB 130 TERIGCGHFE--TLQGVENANHLVCNYPGNVYKGRKPYOEGTGPCSCPLGY--SC 185
QY 230 RNNLCYREETYTPKPTDEMEYMTWDELKSAAMASQICWEHPTSLV 256
DB 186 ENSLC--EPMENPERKADSPRVTREVP 210

Search completed: May 25, 2002, 14:38:41
Job time: 661 sec

copy b.1 + 2

GenCore version 4.5
 Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2002, 10:30:04 ; Search time 2776.47 Seconds

(without alignments)
 11337.819 Million cell updates/sec

Title: US-09-667-380A-1
 1491

Perfect score: 1 atgagctgcgtcctcgtgtgtg.....ggatcttctgtcagcagcag 1491

Sequence: 1 atgagctgcgtcctcgtgtgtg.....ggatcttctgtcagcagcag 1491

Scoring table: IDENTITY_NUC
 Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues 3595312

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database :
 1: gb.ba:*
 2: gb.htg:*
 3: gb.in:*
 4: gb.om:*
 5: gb.ov:*
 6: gb.pat:*
 7: gb.ph:*
 8: gb.pl:*
 9: gb.pr:*
 10: gb.ro:*
 11: gb.sts:*
 12: gb.sy:*
 13: gb.un:*
 14: gb.vi:*
 15: em.da:*
 16: em.fun:*
 17: em.hum:*
 18: em.in:*
 19: em.jny:*
 20: em.om:*
 21: em.or:*
 22: em.ov:*
 23: em.pat:*
 24: em.ph:*
 25: em.pl:*
 26: em.ro:*
 27: em.sts:*
 28: em.un:*
 29: em.vi:*
 30: em.htg.hum:*
 31: em.htg.in:*
 32: em.htg.other:*
 33: em.htgo.in:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query
 No. Score Match length DB ID Description

Result	Score	Match length	DB ID	Description
1	1491	100.0	1491	AX101173 Sequence
2	1491	100.0	2272	AX101175 Sequence
3	1487.8	99.8	4574	AX086850 Sequence
4	1487.8	99.8	4574	AX086850 Sequence
5	1487.8	99.8	4877	AX285067 Sequence
6	1487.8	99.8	4877	AX285067 Sequence
7	1487.8	99.8	4877	AX285067 Sequence
8	1486.2	99.7	2400	AX235369 Sequence
9	1474.8	98.9	2403	AX235373 Sequence
10	1452.6	97.4	2412	AX235375 Sequence
11	1393.8	93.5	1669	AX235371 Sequence
12	1286.2	86.3	2305	AX235367 Sequence
13	1109.6	74.4	1690	AX027395 Sequence
14	974.2	65.3	3052	AF109674 Sequence
15	484.4	32.5	2664	AX250094 Sequence
16	484.4	32.5	2664	AX250094 Sequence
17	484.4	32.5	2667	AF329197 Sequence
18	484.4	32.5	2962	BC020514 Sequence
19	458.6	30.8	2543	AF329198 Sequence
20	441.8	29.6	2382	AF329199 Sequence
21	362	24.3	857	BC007689 Sequence
22	269.4	18.1	1498	AX250025 Sequence
23	240	16.1	174714	AC025280 Sequence
24	238.4	16.0	1365	AF329196 Sequence
25	238.4	15.7	1394	AF329195 Sequence
26	218.6	14.7	594	E09937 Sequence
27	218.6	14.7	1440	D45027 Sequence
28	146	9.8	181161	AC092347 Sequence
29	146	9.8	181161	AC092347 Sequence
30	126.4	8.5	128	AX318722 Sequence
31	89.2	6.0	67263	AC101056 Sequence
32	79.8	5.4	1877	AX136261 Sequence
33	75.4	5.1	2026	BC022399 Sequence
34	75.4	5.1	720	AX136546 Sequence
35	74.6	5.0	970	AR026670 Sequence
36	74.6	5.0	970	AR026670 Sequence
37	73	4.9	184558	AC020570 Sequence
38	73	4.9	185123	AC100782 Sequence
39	69.8	4.7	122	AY033439 Sequence
40	68.8	4.6	181267	AL627168 Sequence
41	68.6	4.6	167819	AC103528 Sequence
42	65.2	4.4	160128	AC093982 Sequence
43	64.2	4.3	2415	AB046537 Sequence
44	62.8	4.2	156235	AC084336 Sequence
45	62.8	4.2	175743	AC011632 Sequence

ALIGNMENTS

RESULT 1
 AX101173 1491 bp DNA linear PAT 10-APR-2001
 LOCUS Sequence 1 from Patent WO0121651.
 DEFINITION
 ACCESSION AX101173
 VERSION AX101173.1 GI:13620006
 KEYWORDS
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 (bases 1 to 1491)
 Donoho,G., Turner,C.A., Waller,F., Nehls,M., Friedrich,G.,
 Zambrowicz,B. and Sands,A.T.
 Novel human proteinase inhibitor-like proteins and polynucleotides
 encoding the same
 Patent: WO 0121651-A 1 29-MAR-2001;
 JOURNAL Lexicon Genetics Incorporated (US)
 TITLE Location/Qualifiers
 FEATURES
 source
 1..1491
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

BASE COUNT 368 a 409 c 426 g 288 t

QY 121 aacgaatctactcccggttcgagagccatccccagggagagagagagatctc 180
 DB 336 AACGATCTCAGTCCCGGCTCCGAGAGCCATCCCAAGGAGAGAGATCTCC 395
 QY 181 atgctgcaacaagaatctcgaggcgagtgacgctcaagctccaagaatgagatc 240
 DB 396 atgctgcaacaagaatctcgaggcgagtgacgctcaagctccaagaatgagatc 455
 QY 241 acctggatgagaaactggaagatctgctgagcggtggcgagtcagtcagtc 300
 DB 456 ACCTGGATGAGAACTGGAGAGTCTGCTGAGCGCTGGGCGAGTCAGTCAGTC 515
 QY 301 caagggcccaagtcgtgtgtgtccatcgagaaacttgaggcgctacgtggcgag 360
 DB 516 CAGGGGCCAAGTCGTGTGTGTCCATCGGCGAAGCTGGGCGCTACTGGGCGAG 575
 QY 361 taccgtctccgggttcacatgtagctgctgtagtcagaggtggaagactacac 420
 DB 576 TACGCTCCGGGTTCCATGTCAGTCTGTGATGAGAGGTGAAGACTACACTAC 635
 QY 421 ccctaccgagagtgcaacccctgtgtccagagaggtgtctggggtctatgtcag 480
 DB 636 CCTTACCCGAGCGAGTCGAAACCCCTGTGTCTGAGAGAGTGTCTGGGCTATGTCTACG 695
 QY 481 cactacaacagatgttggtggccaccacaacagatcggtgtgtgtgtaacactgc 540
 DB 696 CACTACACAGATGTTGGGCTCCACCAACAGATCGTGTGTGTGTGAACACTGCG 755
 QY 541 cggagatgactgtctgggagagatttgggagagcggtctacttgtctgcaattat 600
 DB 756 CGGAGATGACTGCTGGGAGAAATTGGAGAACCGGCTCTACTTGTCTCAATTAT 815
 QY 601 tctccaaagggaactgtgtgagagagcccttaagaatgtggcggtctctgag 660
 DB 816 TCTCCAAAGGGAAGTGTGATGGAGAGCCCTTACAGATGTGGCGGCTCTCTGAG 875
 QY 661 tggccaccagctatgtaggagcagctgtaggaacaacttgttaccagagaagaactac 720
 DB 876 TGCCACCAAGCTATGAGAGCAGCTGCAAGAACACTGTGTACGAGAAAGAACTAC 935
 QY 721 actccaaacactgaacgagagatgaaatgaggttgaagagcgtctccatcttgaaga 780
 DB 936 ACTCCAAACCTGAACGAGAGATGATGAGTGTGAAGAGGCTCCATCTCTGAAGAA 995
 QY 781 aacatgttgggtccaaacgaggtgtagagaccacaagcccaagaacactctgcg 840
 DB 996 AACATGTGTGGTCCAAACGAGGTGATGAGACCCCAAGCCCAAGAAACCTCTGCG 1055
 QY 841 gtcaactacatgacccaagtcgtcagatgtagacccaagaatgaaagagtgcaagg 900
 DB 1056 GTCAACTACATGACCCAAAGTGTGATGATGACACCAAGATGAAGAGAGTGCAGAGG 1115
 QY 901 tccaggtgtaacaggttaacagtgcccgagcggtctgtaacccaagggcgaaatctt 960
 DB 1116 TCCAGGTGTAACAGGTACAGTGCACAGAGGCTGCTGAACCAACAGGCAAAATCTTT 1175
 QY 961 ggaatctgtctatgaaagctgtctagcatatgcccgcgcgcacatccactaagggatc 1020
 DB 1176 GGAATCTGTCTATGAAGAGTGTCTATGATGATGAGGCGCGGCGCATCTACGAGATC 1235
 QY 1021 ctgtagtgaaggaagcgtgtgagatacaacgaaggaaggaaggtcccttctgtg 1080
 DB 1236 CTGAGTGAAGGAGGCGCTGTGTGATATCACAGGAGAGGAGGATCCCTTCTCTGTG 1295
 QY 1081 aagctgagagagagcggtcagtcctcctcagcaataacaactccagctcatctg 1140
 DB 1296 AAGCTGAGAGACAGGCGGTGAGTCTCTCAGCAAAATCAAACTTCAAGCTCATCTATG 1355
 QY 1141 gtgtcaaaagtgaagtgagagatttggactgtgtaacgagacgttgcagctgtgccg 1200
 DB 1356 GTGTCAAAAGTGAAGTGAAGATTGGACGTGTACAGACGCTGTGCTGAGCTGTGCCG 1415
 QY 1201 ttgaaagcagcaactcactgcccagaatcctatgttcggacactgtgaaagagaa 1260

DB 1416 TTTGAAAGCCAGCACTCACTGCCCCAAGATCATTTGCTCGGACACTCAAGAGAA 1475
 QY 1261 ccttcactactggttcggtgttggaaaccaactcatgcaagtaactaaatcagtc 1320
 DB 1476 CTTCTCACTGTTGGGTCCGGTTTGGAAACCAATCTATGAGTACTCAAGCATCTGC 1535
 QY 1321 aagaagctgtgcaagcgagagatcatcagcaacagagatgaggggtgagcgtgag 1380
 DB 1536 AAGACAGCTGTGCAAGCGGAGTCAATGCAACAGAGAGTGGGGTGAAGTGAAGTATG 1595
 QY 1381 cccgtgataaagaagaactacgtgtggtcgtcaggaatgagttcagtcgaagc 1440
 DB 1596 CCCGTGATTAAGAAAGAACCTACAGTGGGCTCGCTCAGAGATGAGTCACTGGAAGC 1655
 QY 1441 ctggagactcctcgagatgagaaagccttcagatcttctgtcagagag 1491
 DB 1656 CTGGGACTCTCGGATGAGAAAGGCTTCCGAGATCTTGTGTGACAGCAG 1706

RESULT 3

LOCUS AX086850 4574 bp DNA linear PAT 09-MAR-2001
 DEFINITION Sequence 802 from Patent WO0112659.
 ACCESSION AX086850
 VERSION AX086850.1 GI:13276048
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE

1 (bases 1 to 4574)
 AUTHORS Wiemann, S.
 TITLE Human dna sequences
 JOURNAL Patient: WO 0112659-A 802 22-FEB-2001;
 German Human Genome Project (DE)

FEATURES

source location/Qualifiers
 1..4574
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 BASE COUNT 1105 a 1166 c 1158 g 1145 t
 ORIGIN

Query Match 99.8%; Score 1487.8; DB 6; Length 4574;
 Best local Similarity 99.9%; Pred. No. 0;
 Matches 1489; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atgagctggtctctgggtgtgtgtatcccttggggctgtcttccgtgtcgatcc 60
 DB 205 ATGAGCTGCTCTCTGGGTGTGTGTATCCCTTGGGCTGTCTGTGTCTGGGATCC 264
 QY 61 caaggtactcctctgccaagctcactctcttagaagagctgtcagaatacagaac 120
 DB 265 CAAGGTACTCTCTGCCCAAGCTCACTCTCTTAGAGAGCTGTCAAGAAATACAGCAC 324
 QY 121 aacgagcttaactccgggttcgcagagacatccccaaggagaaaggaagatcttc 180
 DB 325 AACGAGTCTCACTCCGCGTCCGACAGCCATCCCAAGGAGAGAGAGATCTCTC 384
 QY 181 atgtgcaacaagaactctggggccaaggtgcaagcctccaagcctccaacatgagatc 240
 DB 385 ATGTGCAACAAGCTCTGGGCGCAGGTGCAAGCTCAAGCTCAAGATGAGTACAG 444
 QY 241 acctggatgagaaactggaagatctgtctgacagctgtggcgagtaatgcatctggag 300
 DB 445 ACCTGGATGAGAACTGGAGAGTGTGTGACAGCGTGGGCGCAGTCAATCTTGGAG 504
 QY 301 caggggcccacagctgtcgtgtgtgtcaatgggcagaagaactggcgctcactgtggag 360
 DB 505 CAGGGGCCCAACAGCTGTGTGTGTGTCAATGGGCGAAGCTGGGCGCTCACTGGGCGAG 564
 QY 361 tatgctctccgggttccatgtgcaatctgtatgtagcaaggtlgaagagactaacactac 420

```

Db 565 TATCGCTCCGCGGCTTCATGTCAGTCTGATGACGAGGAGACATACACCTAC 624
QY 421 cccctaccgagcgagtgcaacccccgtgtccagagagtgctcggggcctatgtcagc 480
Db 625 CCGTACCCGAGCGAGTCCACACCCCTGGTGTCCAGAGAGTCTCGGGCCCTATGTGACG 684
QY 481 cactacacacagatagtttggccacacacacacacagatcgtgtgtgttaaacctcgc 540
Db 685 CACTACACACAGATAGTTTGGGCCACACCAACAAAGATCGTTTGTCTGTAACACCTGC 744
QY 541 cggagagatgactgtctgggggagagtttgggagaaacggtctactctgtctgaattat 600
Db 745 CGGAAGATGACTGTCTGGGAGAAAGTTGGAGAACCGGCTCTACTTGTCTGCAATTAAT 804
QY 601 tctccaaaggaggaactgtagtggagaaagccccctacagaatggccggccctcgtcag 660
Db 805 TCTCCAAAGGGGAACTGGATTGGAGAACCCCTCCACAAAGATGGCCGCGCTGCTGAG 864
QY 661 tggccaccccgactatggagcagctgcaggaacaacttgtttaccgagagaagaacctac 720
Db 865 TGCCCAACCCAGCTATGAGGAGCAGCTGCAGAACACTTGTGTACGAGAGAAACCTAC 924
QY 721 actccaaactgaaacgagagatgaatgaagtgtgaaacggtcccatctcgtgaaga 780
Db 925 ACTCCAAACCTGTAACGAGAGAGATGAATGAGTGAAGGCGGCTCCCATTTCTGAGAA 984
QY 781 aacacatgttggtctccacacgaggtgatagagaccccaagcccaagaagaacctctgc 840
Db 985 AACCATGTTGGCTCCCAACGAGGAGTGTAGAGACCCACCAAGCCCAAGAAACCTCTGC 1044
QY 841 gttcaactacatgacccaagtcgtcagatgtgacacacagatgaagcaggtgcaagg 900
Db 1045 GTCAACTACTATGACCCCAAGGCTGTCAATGTGACACCAAGATGAGAGAGAGTCAAGG 1104
QY 901 tccacgtgttaacagtgacagtgatccacagcaggtgtcctaacaacagcgagatcctt 960
Db 1105 TCCACGCTGTAACAGTATGACAGTGTCCAGAGGCTGCTGTAACACCAAGCGAGATCTTT 1164
QY 961 ggaagctgttctatgaagctcgtctagacatagtcgcgcgcgcacatccactacgggac 1020
Db 1165 GGAAGCTGTGTTATGAAGCTGTGCTACATATGCGCGCCGCAATCCACTACGGGATC 1224
QY 1021 ctggatgacaaaggagcctgtgtgtatcatcacaagaaaggagagagtcctcttcgtg 1080
Db 1225 CTGGATGACAAAGGAGCGCTGTGTGATATCACAGAGAGAGAGAGTCCCTTCTTCGTG 1284
QY 1081 aagctcgaagacagcggtgtcagtcctcagcaataacacactccagctcattcatg 1140
Db 1285 AAGCTCGAAGACACGCGCTGTGACGTCCTCACCAATATCAACCTTCCAGCTCATTTCA 1344
QY 1141 gtgtcaaaagtgaagtgagagatttgtagctgtctacagacggtgtgtcagctgtgcgg 1200
Db 1345 GTGTCAAAAGTGAAGTGTGAGATTGTGACTGTACACGACCGCTGTGCTGAGCTGTGCCG 1404
QY 1201 ttgtgaaagccagcactactcgtcccaagaatcctatgtccgggacactgtcaaaagaa 1260
Db 1405 TTTGAAAAGCCAGCACTCACTGCCCAAGAAATCCATTTGTCGGGCACTGTCAAAAGCAA 1464
QY 1261 ccttctactcgtgtcgcggtgtttgtgaaaccaacatctatgtcagataccctcagctcgc 1320
Db 1465 CTTTCTCTACTGGGCTCCGCTGTTTGGAAACCAACATCTATGAGATACCTCAAGCATCTGC 1524
QY 1321 aagaagcgtgtgacagcgaggagtcacagcaagagatgggggtgaggtgagcgtgtagt 1380
Db 1525 AAGACAGCCGTCACGCGGAGTCTATCAGCAACGAGAGTGGGGGTGACGTGAGCTGATG 1584
QY 1381 cccgtgtgataaaagaagactacagtggtcgtcagcaagaatgagttcagttcgtgaaagc 1440
Db 1585 CCGGTGATATAAAAGAAAGAACTACAGTGGGCTCGCTCAAGAGATGAGTTCAATCTGANAAGC 1644
QY 1441 ctggagactcctcggagatgagaaagcctcgcgagctcttctgtcagcag 1491

```

```

Db 1645 CTGGGAGCTCTCTCGGAGTGAAGAGCCCTTCGAGATCTTGTCTGTCAGCAG 1695
RESULT 4
HSM801829 4574 bp mRNA linear PRI 03-AUG-2001
LOCUS HSM801829 Homo sapiens mRNA: cDNA DKFp434B044 (from clone DKFp434B044);
DEFINITION complete cds.
ACCESSION AL136861
VERSION AL136861.1 GI:12053226
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 4574)
AUTHORS Wambutt,R., Heuber,D., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
TITLE Direct Submission
JOURNAL Submitted (15-JUN-2001) MIPS, Am Klopferspitz 18a, D-82152
Martinsried, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
consortium of the German Genome Project.
This clone (DKFp434B044) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heuberweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at http://mips.gsf.de/proj/cdna/.
FEATURES
source
1..4574
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="333.4 kb from top of Chr16 linkage group"
/clone="DKFp434B044"
/rissue="testis"
/clone_id="434 (synonym: htes3). Vector pSport1; host
DH10B; sites NotI + SalI"
/dev_stage="adult"
205..1698
/gene="DKFp434B044"
205..1698
/gene="DKFp434B044"
/note="strong similarity to trypsin inhibitor."
/codon_start=1
/product="hypothetical protein"
/protein_id="CA866795.1"
/db_xref="GI:12053227"
/db_xref="SPTREMBL:O9H0B8"
/translation="MSCVLGCVIPILFLFVCGSOGYLLPNTLTLEELISKYONHESH
SVRRRAIPREDKEELIMLNKLRGGOVOPASNNVETMPDELKSAAMASOCIEHG
PTSLVLSIGONIGAHMGRYSRSEPHQSTYDEKDTTTPSECNWCPERSGPMCT
SESPSTYSGSCRNKLCTYRETYTPKPEDEMEVAVFVCYSPKGNMIGAPAKNRPK
KTSVAVMTQVNAQCDTKMKDRKSGTSCNAYOCQACGLNKAIFGLFPESSSICRA
ALHYGLIDKGGGLVDITRNGKGFYFKSRHGVSSIKYKPSSEFMVGVKXKODIDCY
TVAOGLPEKPEPATHCPRILHCPAHCKDEPSYVAAPVPCNTIYADTSICKTANHAGVIS
NESGQVDVMPYDKKTYGSLRNGVSSLSLTPDGRKAFRIFAVQ."
4540..4545
polya_signal
BASE COUNT 1105 a 1166 c 1158 g 1145 t
ORIGIN
Query Match 99.8%; Score 1487.8; DB 9; Length 4574;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1489; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 atgagctgcctctgggtgtgtcaccctctgggagctgtctcgtgtctgggagatcc 60
Db 205 ATGAGCTGCGTCTGGGTGTGTCAATCCCTTGAGCTGCTGCTGCTGCGGATCC 264
QY 61 caaggtactcctcgtcccaagctactctcttagagagagctcagcaataacagcac 120

```



```

Db 265 CAAGGCTACTCTGCGCAACGTCACCTCTTATGAGAGAGTGTCTGACGAATACAGCAC 324
Qy 121 aacgagatcactccgggtccgacagacatccccaaggaggaacagaagatccctc 180
Db 325 AACGAGTCTACCTCCGGGTCGCCAGAGCCATCCAGGAGACAAAGAGGAGATCTTC 384
Qy 181 atctctcaacaagaagctctgggccaagtgacagccccaagctcgaactgagtaacatg 240
Db 385 ATGTGTGACAAACAAAGCTTCGGGGCCAGGTGCACCTCAGCCCTCCAAACATGAGTACATG 444
Qy 241 acctggatgacgaactggaagatctgctgcagcgtyggcagtycaatgacatctggag 300
Db 445 ACCGTGGATGACGAACCTGAGAGATGCTGCTGCGACGCGTGGGCCAGTCACTGCTGGAG 504
Qy 301 caagggcccaacagctcgtcgtgtccatctggcagaacactgggcccctcaactgggacg 360
Db 505 CACGGGCCCAACGATGCTGCTGCTCCATGCGCAGAACTGGGCGCTCTCACTGGGCGAG 564
Qy 361 tatcgctccggggttccatgtgcaatctctgtatgacagagtgaaagatacaactac 420
Db 565 TATGGCTCTCGGGGTTCCATGTGCATCTCTGTATGACAGGTGAAGGATACACTAC 624
Qy 421 cccctaccgagcgagtgcaaacccctggtytccagagagtgctcggggacctatgtcag 480
Db 625 CCTTACCCGACGAGTGCACACCCCTGGTGTCCAGAGAGTGTCTGGGCGCTATGTGCAGC 684
Qy 481 cactacaacagatgaattgtggccacaacaagaatcggtgtgtgtgtaaacatctgc 540
Db 685 CACTACACACAGATGATGTTGGGCCACCAACACAGATGGTGTGTGTGTGTAACACTGTC 744
Qy 541 cgaagatgactgctcgtggggaagatttgggagacgcggttctaattgtctgcaattat 600
Db 745 CGGAAGATGATGCTGTGGGGGAGGTTGGAGAACGGCGCTACTTGTGTGCAATATAT 804
Qy 601 tctccaaagggaactggaattggagaaagccctcaagaatggcgccgtctgtctgag 660
Db 805 TCTCCAAAGGGGAACTGGATTTGGAGAAACCCCTTCAAAAGATGGCCGCTGTCTGTAG 864
Qy 661 tggccaccagactatggagagcagctgcaggaacaactgtgttllcccgaaagaacactac 720
Db 865 TGCCACCACCACTATGGAGGACGCTGCAGAAACAACCTGTGTACCGAAGAAACCTAC 924
Qy 721 actccaaactgaaacgagacgagatgaaatgagtgaaacgagctccactctgaagaa 780
Db 925 ACTCCAAACCTGAAACGAGACGAGATGATGAGTGGGAAAGGCTCCATTCCTGAAATA 984
Qy 781 aacctgtttgtctcaacgagaggtgtagaagcccaagcccaagaagaactctgcg 840
Db 985 AACCAATGTTGGCTCCAAACCGAGGCTGATGAGACCCACCAAGGCCCAAGAAACCTCTGCG 1044
Qy 841 gtcaactacatgaccgaagctcgtcagatgtgaacacaaagtgaaggaagtgcaaaagg 900
Db 1045 GTCAACTACATGACCAAGTGTCTGAGATGTGACACCAAGATGAAGACAGGTGCAAAAGG 1104
Qy 901 tccacgltgaacaggttaccagtgcccagcagagctcgtcgaacacaaaggcgaatctt 960
Db 1105 TCCACGTTGTAACAGGTACAGTGCACAGCAGGCTGCTGAAACCAAGCGGAAGATCTTT 1164
Qy 961 ggaagctgtttatgaagaagctcgtctagaatgacgcgcgcgcacatccacagggatc 1020
Db 1165 GGAAGCTGTCTTCTGAAAGCTGCTTACATATGCGCGCGCCATCCACTACGGGATC 1224
Qy 1021 ctgagatgaagaagagccttgglygatalcaccagaagacggaaggttccctcttctgt 1080
Db 1225 CTGAGATGAAGAAGGAGCCCTGGGATATCACAGGAAGGAGGTCCCTTCTTCTG 1284
Qy 1081 aagctgagagacagcggtgtgacgtccctcagcaaaataaaactccagctcattcatg 1140
Db 1285 AAGTCTGAGAGACACGCGGTGACGTCCCTCAGCAAAATTAACCTTCAGCTCATCATG 1344
Qy 1141 gtctcaaaagtgaagtgcaagattggagctgtacacagcgttgcagctgtgtccg 1200

```

```

Db 1345 GTGTCAAAAGTGAAGTCCAGGATTTGAGCTGCTACACAGCCGTTGCTGACCTGTGCCG 1404
Qy 1201 ttgaaagccagcaactcacttcccaagaatccatgtgtccgacactgcaaaagaa 1260
Db 1405 TTTGAAAGCCAGCAACCTACAGTCCCAAGAAATCCATFTCTCCGGCACACTGCAAAAGACGA 1464
Qy 1261 ccttcactaggtcgcgtgtgttggaaacaaatctatgtgaatgaactgaatcgtgc 1320
Db 1465 CTTTCTACTGGGCTCCGCTTTTGGAAACCAATCTATGACGATGCTTCAAGCACTTCTGC 1524
Qy 1321 aagacagctgtgacagcggaatccatcagcaagagatgggggtgagctgagctgag 1380
Db 1525 AAGACACCGCTGACGCGGGAGTCTATCAGCAACGAGATGGGGGTGACGTGGAGTATG 1584
Qy 1381 cccgtgataaaagaagaactcagtgaggctgcctcaggaatgagtgatcagtgaaagc 1440
Db 1585 CCGTGATTAAGAAAGAACACTACGTGAGGCTGCTCAGGAATGAGATTCACTGTAAGC 1644
Qy 1441 ctggggaactcctcggatgagaaagccttcggatccttgcgtgacagcag 1491
Db 1645 CTGGGACTCTCGGGATGAAAGGCTTCGGAATCTTGTGTGACAGGCGAG 1695

RESULT 5
AX285067/c 4877 bp DNA linear PAT 20-NOV-2001
LOCUS Sequence 872 from Patent WO01/9556.
DEFINITION AX285067
ACCESSION AX285067
VERSION AX285067.1 GI:17045755
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (sites)
AUTHORS Lilje,J., Brown,J.L., Bolt,A. and van Hufel,C.
TITLE Novel genes, compositions and methods for the identification,
assessment, prevention, and therapy of human cancers
JOURNAL Patent: WO 01/9556-A 872 25-OCT-2001;
Millennium Predictive Medicine, Inc. (US)
FEATURES
source
BASE COUNT 1183 a 1201 c 1236 g 1257 t
ORIGIN
Query Match 99.8%; Score 1487.8; DB 6; Length 4877;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1489; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 atgagctgcgtcctcgtggtgtgtcattcccttgaggctcgtgtctcgtcgtgcagtc 60
Db 4657 ATGAGCTGCGCTCGTGGGTGGTGTATCCCTTGGGGCTGTGTCTCGTCTCGGATCC 4598
Qy 61 caaggttaactccctggcccaaggtcactctttagaaggagctgctcagcaataaccagcac 120
Db 4597 CAAGGCTACCTCTGCCCCAACGTCACTCTTTAGAGAGAGTGTCTGACAAATACAGACAC 4538
Qy 121 aagagctcactccgggtgtccgacagacatccccaaggaggaacaggaagatccctc 180
Db 4537 AACAGGTCTACTCCCGGGTCCGACAGACCATCCCAAGGAGGACAAAGAGAGATCTCTC 4478
Qy 181 atgtctcaacaagaactcctgggccaaggtgacgctcaagctcacaatagagatcatg 240
Db 4477 ATGTGTCAACCAACCTTTCGGGGCCAGGTGACGCTCAAGGCTCCCAATGAGATGATCATG 4418
Qy 241 acctggatgacgaactggaagatctgtcagcgttggcgcagtgcaatcagtgcatctggag 300
Db 4417 ACCTGGATGACGAACCTGGAAGATCTCTGACAGGTGGGCGACAGTGCATGTGGAG 4358
Qy 301 caagggcccaacagctcgtgtgtcactcggcagaacactggcgctcactcagggcag 360

```

Db	3277	CCCGTGGATATAAAAAGAAAGACCTACGTGGGGCTCGCTCAGGAATGGAGTTCACTGATGAAAGC	3218
OY	1441	ctgggagctctctcggaatgaaagcctccgaatctcttgctgacgagcag	1491
Db	3217	CTGGGCACTCTCGGATGGAAAGCGCTTCGCGATCTTTGCTGTGACGACG	3167
RESULT	6		
AX285068/c	AX285068	4877 bp	DNA
LOCUS	Sequence	873 from Patent WO0179556.	linear
DEFINITION	AX285068		
ACCESSION	AX285068.1	GI:17045756	
VERSION			
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (sites)		
JOURNAL	Lillie,J., Brown,J.L., Bolt,A. and van Huffel,C.		
FEATURES	Novel genes, compositions and methods for the identification, assessment, prevention, and therapy of human cancers		
source	Patent: WO 0179556-A 873 25-OCT-2001;		
	Millennium Predictive Medicine, Inc. (US)		
	Location/Qualifiers		
	1..4877		
BASE COUNT	1183 a	1201 c	1236 g 1257 t
ORIGIN			
Query Match	99.8%; Score 1487.8; DB 6; Length 4877;		
Best Local Similarity	99.9%; Pred. No. 0;		
Matches 1489; Conservative	0; Mismatches 2; Indels 0; Gaps 0;		
OY	1	atgaagtcgctcctgggtggtgctgcatcccttggggctcgtcttcctgctcggaatcc	60
Db	4657	ATGAGCTCGTCTCTGGTGTGTGTCATCCCTTGGGGCTGCTTCTTCGTGTCGGGATCC	4598
OY	61	caaggctactctctctgccaagctactctcttaagaggagctgctcaagaatatacagaac	120
Db	4597	CANAGCTACTCTCTGCGCAACGTCACTCTTTAGAGGAGCTGCTCAGCAAAATACAGCAC	4538
OY	121	aacgagctctcaatcccggttcgcaagacatcccccaaggaggaaagaaagagatccctc	180
Db	4537	AACGAGTCTCACTCCCGGGTCCGCAAGCCATCCCAAGGAGAGATAAGAGAGATCTCTC	4478
OY	181	atgctgcacaacaagcttcgggccaagtgaaagctcagagctccaacatggagtacatg	240
Db	4477	ATGCTGCACAACCAACTTCGGGGCCAGGTGACGCTTCAGCGCTCCAACTGAGATCATG	4418
OY	241	acctggatgagcaactggagaatgctctctcagagtggtggccaatgacatctggag	300
Db	4417	ACCTGGATGAGCAACTGGAGAAAGTCTCTGTAAGGTGAGGCGACGTGATGTGGAG	4358
OY	301	caagggccaccagctctgctggtgtcaatcgggcaagaacctgggacgtcaactgggcaag	360
Db	4357	CACGGGCCACACAGTCTGCTGGTGCATCGGGCAAGAACTGGGGCGTCACTGGGGCAGG	4298
OY	361	tatgctctcggggtgttcatagttcagatccctggtatgaagaagtgaagactacaactac	420
Db	4297	TATGCTCTCCGGGGGTTCATGTGCAAGTCCGATATGACGAGGTGAAGGACTACACCTTAC	4238
OY	421	ccctaccgagcgaagtgaacccctgtgtccaaagagagtgctctggggcctatgtgaag	480
Db	4237	CCCTACCAGACGAGTGAACCCCTGGGTGTCAAAAGAGTGCTCGGGCCATTGTGACG	4178
OY	481	cactacacacagaatgatttgggccaccacaacaagaatcggtgtgtgtgtgaacactgc	540
Db	4177	CACCTACACACAGAAATGTTGGGCCACACCAACCAAGATCGGTGTGCTGTGAACACCTGC	4118
OY	541	cgagaagatgactgtccttgqggaagaagtttggagaaanacnrgtctactttttctcaatttc	600

[illegible]

Db 1046 GCGGCACTACATGACCCAGCTGATGTGACACCAAGATGACAGGAGAGAGTGCAGAA 1105
 Qy 898 GGTGCACTGTAACAGTACCAGTCCAGCAAGGCTGCTGCAACCAAGAGAGAGTGCAGAA 957
 Db 1106 GGTGCACTGTAACAGTACCAGTCCAGCAAGGCTGCTGCAACCAAGAGAGATC 1165
 Qy 958 GGTGCACTGTAACAGTACCAGTCCAGCAAGGCTGCTGCAACCAAGAGATC 1017
 Db 1166 GGTGCACTGTAACAGTACCAGTCCAGCAAGGCTGCTGCAACCAAGAGATC 1225
 Qy 1018 GGTGCACTGTAACAGTACCAGTCCAGCAAGGCTGCTGCAACCAAGAGATC 1077
 Db 1226 GGTGCACTGTAACAGTACCAGTCCAGCAAGGCTGCTGCAACCAAGAGATC 1285
 Qy 1078 GGTGCACTGTAACAGTACCAGTCCAGCAAGGCTGCTGCAACCAAGAGATC 1137
 Db 1286 GGTGCACTGTAACAGTACCAGTCCAGCAAGGCTGCTGCAACCAAGAGATC 1345
 Qy 1138 GGTGCACTGTAACAGTACCAGTCCAGCAAGGCTGCTGCAACCAAGAGATC 1197
 Db 1346 GGTGCACTGTAACAGTACCAGTCCAGCAAGGCTGCTGCAACCAAGAGATC 1405
 Qy 1198 GGTGCACTGTAACAGTACCAGTCCAGCAAGGCTGCTGCAACCAAGAGATC 1257
 Db 1406 GGTGCACTGTAACAGTACCAGTCCAGCAAGGCTGCTGCAACCAAGAGATC 1465
 Qy 1258 GGTGCACTGTAACAGTACCAGTCCAGCAAGGCTGCTGCAACCAAGAGATC 1317
 Db 1466 GGTGCACTGTAACAGTACCAGTCCAGCAAGGCTGCTGCAACCAAGAGATC 1525
 Qy 1318 GGTGCACTGTAACAGTACCAGTCCAGCAAGGCTGCTGCAACCAAGAGATC 1377
 Db 1526 GGTGCACTGTAACAGTACCAGTCCAGCAAGGCTGCTGCAACCAAGAGATC 1585
 Qy 1378 GGTGCACTGTAACAGTACCAGTCCAGCAAGGCTGCTGCAACCAAGAGATC 1437
 Db 1586 GGTGCACTGTAACAGTACCAGTCCAGCAAGGCTGCTGCAACCAAGAGATC 1645
 Qy 1438 GGTGCACTGTAACAGTACCAGTCCAGCAAGGCTGCTGCAACCAAGAGATC 1491
 Db 1646 GGTGCACTGTAACAGTACCAGTCCAGCAAGGCTGCTGCAACCAAGAGATC 1699

RESULT 10
 AX235375
 LOCUS AX235375 2412 bp DNA linear PAT 11-SEP-2001
 DEFINITION Sequence 22 from Patent W00162928.
 ACCESSION AX235375
 VERSION AX235375.1 GI:15593907
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 2412)
 Vernet,C.A., Fernandes,E., Shlunkels,R.A., Macdougall,J. and
 Spaderna,S.K.
 Polypeptides and nucleic acids encoding same
 Patent: WO 0162928-A 22 30-AUG-2001;
 Curagen Corporation (us)
 TITLE
 JOURNAL
 FEATURES
 source
 1. 2412
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /note="unnamed protein product"
 /protein_id="CAC69821.1"
 /db_xref="GI:15593908"
 /translation="MSCYLGIVPLGILFLVRSQGYLLPNVTLLELLSKYOHNEH
 SRVRAIPREDKEILIMLHNLKRGVOPQASNMETWDELKESAAAMASOCIMERG
 PTGLIVSIGONI GAHMGRIYRSPFHVOSWDEVKDYNY PYPSECPNCPGRCGPMTCT

Query Match 97.4%; Score 1452.6; DB 6; Length 2412;
 Best local Similarity 98.9%; Pred No. 0;
 Matches 1487; Conservative 0; Mismatches 4; Indels 12; Gaps 2;
 BASE COUNT 560 a 649 c 688 g 515 t
 ORIGIN
 HVTQVOTVWATTTNKIGCAVNTCRMVWGEVWENAVYVONYSPPKNGMIGEPYKN
 GRPCESECPVSGSCRNMI,CYREETVTKPEIDEMNEVEVTAIPIDENHVMLOPRMR
 TPKRTSAVNTVQVRODTEKMDKSKSTCNRYOCPPGICNHRKIGTSLFESSSS
 ICRAAIIHGLDDKGLVDITPNKGVPEFVSESRGVOSLSKYRSPSMYSKYKVD
 LDCYVLOLCEFEKPATHCPRIHCPANCXDEPSMAYPEFTNTYADTSICRTAVHA
 GVISNESGDDVDPVDDKKRTYVGLSRNVOSSESIGTPRDKAFRIPAVRO"
 Qy 1 atgagctgcgttcctgggtgtgtcacccttgggtgtgtcctgtgtcgtgcgtgc 60
 Db 206 atgagctgcgttcctgggtgtgtcacccttgggtgtgtcctgtgtcgtgcgtgc 265
 Qy 61 caaggtacctctctgcccagctcctcttaaggaggtgtgtcgtcgaataaccagcac 120
 Db 266 caaggtacctctctgcccagctcctcttaaggaggtgtgtcgtcgaataaccagcac 325
 Qy 121 aacgagctcactcccggtgtccgagagccatcccccagagagagagagagagagag 180
 Db 326 aacgagctcactcccggtgtccgagagccatcccccagagagagagagagagagag 385
 Qy 181 atgctgcaacaacaacttcggggcagaggtgtcagctcagctccacaatgagatgac 240
 Db 386 atgctgcaacaacaacttcggggcagaggtgtcagctcagctccacaatgagatgac 445
 Qy 241 acctggatgaagaaactgtgagagctgtcgtcagaggtgtgagagagagagagagag 300
 Db 446 acctggatgaagaaactgtgagagctgtcgtcagaggtgtgagagagagagagagag 505
 Qy 301 caagggcccaacagctgtgtgtgtcactcggagaaactcgggcgtcactcgggcagag 360
 Db 506 caagggcccaacagctgtgtgtgtcactcggagaaactcgggcgtcactcgggcagag 565
 Qy 361 tatgccttcgggggttcctatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 420
 Db 566 tatgccttcgggggttcctatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 625
 Qy 421 ccttaccgagagaggt 480
 Db 626 ccttaccgagagaggt 685
 Qy 481 cactac 531
 Db 686 cactac 745
 Qy 532 aacacctgcggaagatgactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 591
 Db 746 aacacctgcggaagatgactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 805
 Qy 592 tgcgaattatctccaaa---ggggaactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 648
 Db 806 tgcgaattatctccaaa---ggggaactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 865
 Qy 649 ccttgcctgt 708
 Db 866 ccttgcctgt 925
 Qy 709 gaagaaacttaccac 768
 Db 926 gaagaaacttaccac 985
 Qy 769 attcctgt 828
 Db 986 attcctgt 1045
 Qy 829 aaaaactgt 888
 Db 1046 aaaaactgt 1105

QY 889 aggtgcaagagtcacagtcgtacacagtcgccagcgctgctgtaacacacag 948
 |||||
 Db 1106 AGGTGCAAGAGTCCACGCTGTAACAGGTACAGTGCACGAGGCTGCTGAACCAAG 1165
 |||||
 QY 949 ggcgaatcttggaaagtcgtctctatgaagctgctagacatatgcgcgcgcac 1008
 |||||
 Db 1166 GCGAAGATCTTGGAGTCTGTTCTATGAAGCTCTGATGACATAGCCGCCGCCATC 1225
 |||||
 QY 1009 cactacggatctctgatatgacaaggagcctggatgatatcaacaggaacggagtc 1068
 |||||
 Db 1226 CACTACGGGATCTCTGATGACAGGAGGCTGTGATATCACAGGAACGGAGATC 1285
 |||||
 QY 1069 cccctctctgtaagtcgtgagaaacgcggtgtaacccctcagaaatacaacccctc 1128
 |||||
 Db 1286 CCGCTTCTCTGTAAGTCTGAGAACAGCGGCTGAGTCCCTCAGCAATACCAACCTTCC 1345
 |||||
 QY 1129 agctcattcatggttcaaaagtgaagtcagagatttgactgtctacacgacgttgc 1188
 |||||
 Db 1346 AGCTCATTCATGCTGTAAGTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1405
 |||||
 QY 1189 cagctgtgcccgttgaagaaagcagcaactcactgcccagaatcattgtccggagac 1248
 |||||
 Db 1406 CAGCTGTGCCGTTTGAAGAACCCAGCAACTCACTGCCAAGAAATCAATGTCCGGCAC 1465
 |||||
 QY 1249 tgcgaagaagaaactcctcactggtcgtcgtgttggaaacaaatctatgagatacc 1308
 |||||
 Db 1466 TGCAAAAGACAACTCTCTACTGCGGCTCGGCTGTGGAACCAACATCTGACAGATAC 1525
 |||||
 QY 1309 tcaagcatctgcaagaagcagctgtgacgcgagagatcagcaagagatgggtgac 1368
 |||||
 Db 1526 TCAAGCATCTCTCAAGACAGCCGCTGACGCGAGTCAATGCAAGAGAGTGGGGTAC 1585
 |||||
 QY 1369 gtgagatgtatgcccgtgagataaaagaagacactgagtcgcgcaggaatgagtt 1428
 |||||
 Db 1586 GTGAGCGTATGATCCGCTGATGAAGAAAGACACTACGTGGGCTGCTCAGGAATGAGTT 1645
 |||||
 QY 1429 cagctcgaagacgtgagacactcctcgtgagatgaaagcctcgcgattctgtcgaag 1488
 |||||
 Db 1646 CAGTCTGAAACCTGGGGACCTCTCGGATGGAAGGCTCTCGGATCTGTGTCAGG 1705
 |||||
 QY 1489 cag 1491
 |||||
 Db 1706 CAG 1708

RESULT 11
 AX235371 1669 bp DNA linear PAT 11-SEP-2001
 LOCUS AX235371
 DEFINITION Sequence 18 from Patent WO0162928.
 ACCESSION AX235371
 VERSION AX235371.1 GI:15593903
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 Vernet,C.A., Fernandes,E., Shmkelts,R.A., Macdougall,J. and
 Spaderna,S.K.
 TITLE Polypeptides and nucleic acids encoding same
 JOURNAL Patent: WO 0162928-A 18 30-AUG-2001;
 Curagen Corporation (US)
 FEATURES
 source
 1..1669
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 205..1638
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAC69819.1"
 /db_xref="GI:15593904"
 /translation="MSCVIGVIVPLGLFLVCGSGVLLPNTLLLELLSKYQHNESH

SVRRRAIPREDKEEILMLANKLGQVOPASNSMEYTWDELEKSAANAASOC IWEHG
 PTGLIVSIGNLGKHAMGRYSRPGFHVOSWDEKDYTPYPSDCNCEPRGSGPCT
 HYDIOIVATNLCIGAVNCRKMTVMWEVENVAVYCNTPSPGNMIGENAPYKGRPC
 SOCPYSYSGCRNNLCVREFTYKPKEDNREVENAP IPEENHVMLOPRMPTPK
 KTSSVNMTOYVLCDITKMRCKGSCNDRNOCAGCINHEAK IEGTLFEYSSSICRA
 AIHGYLLDDGGLVDITRNGKVPFFVSEHGVOSLSKYPSSSPVMSKYKVDLDY
 TTVVQLCPFERPAHCPRIHCPACCKDEPSTWMAVFQTNITADTSSICKRAVAVGIS
 NESGSDVDMVMDKRTTCTCPAARAL"
 BASE COUNT 388 a 484 c 476 g 321 t
 ORIGIN
 Query Match 93.5%, Score 1393.8; DB 6; Length 1669;
 Best Local Similarity 99.2%; Pred. No. 0;
 Matches 1401; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
 QY 1 atgagctggtctctggtggtgtatccctctggggtcgtctctgctgctgacac 60
 |||||
 Db 205 ATGAGCTGCTGCTCTGGGTGTGTCTATCCCTTGGGGCTGCTCTGCTGGCGATCC 264
 |||||
 QY 61 caagctacccctctgcccacagtcactctctatgagagctgtctcaagaataaccagac 120
 |||||
 Db 265 CAAGCTACCTCTCTGCGCAAGCTCACTCTTGAAGAGACTGCTCAGCAATACAGCAC 324
 |||||
 QY 121 aacgagctcactcccggttcgcagagacatcccgagagagagagagagatctc 180
 |||||
 Db 325 AACGAGTCTCAGTCCCGGTTCGCGAGAGCCATCCCGAGGAGACAGAGAGATCTTC 384
 |||||
 QY 181 atgtcacacaagaactctggggccaggtgacagctcagagctccaacatgagatagatg 240
 |||||
 Db 385 ATCTGTCAACAAGAGTGTGGGGCCAGGTGACAGCTCAGGCTCCCAATGAGATACAG 444
 |||||
 QY 241 aactggagatgagaaagtcgtctgctgacgctggggcaagtcagagtcctggagag 300
 |||||
 Db 445 ACCTGGATATCAGCAACTGGAAGTGTGCTGCAAGCTGGAGCAAGTATGCTGGAG 504
 |||||
 QY 301 caggggcccaacagctgtcgtgtgtcactgcagcagaaactcgtggcgtcactcgtggcag 360
 |||||
 Db 505 CAGGGGCCCAACCGGTGTGCTGTCTCATGCGGCAACACTGGGGCTCACTGGGGCAGG 564
 |||||
 QY 361 tctcgtctccgggttccatgtgacagtcgtctgtagcaagatgaaagtaactacacac 420
 |||||
 Db 565 TATCGCTCTCCGGGTTCCATGTGCACTGCTGATGACAGGTGAAGGATCAACCTTAC 624
 |||||
 QY 421 cctcaccgagagagtgcaacccctgtgtccagagagatgtcgcgggacctatgtacag 480
 |||||
 Db 625 CCTTACCCGAGCAGTCAACCCCTGTGTGCAAGAGTGTCTGCGGGCTATGTGTCAGG 684
 |||||
 QY 481 cactacacacagatagtttgggccaacacacacacacacacacacacacacacacac 540
 |||||
 Db 685 CACTACACACAGATAGTTTGGGCCACACACACACACACACACACACACACACACAC 744
 |||||
 QY 541 cgaagatgactgtctggtggaagattgggagaagcgtctacttctgtcgaattat 600
 |||||
 Db 745 CGGAAGATGACTGTCTGGGGAGAGTTGGAGAACCGGCTCTTACTTGTCTCAANTAT 804
 |||||
 QY 601 tctcacaagagagactgattggaagacccctacagaatgtgcgcgcctgtcgtgag 660
 |||||
 Db 805 TCTCCAAAGGGAAGTGTGAGAAAGCCCTACAAAGATGGCCGCGCTGCTCTACG 864
 |||||
 QY 661 tgcacacacagatgtggaagcagctgcagagaacacttggttacccgagaagaacctac 720
 |||||
 Db 865 TGCCACACCACTATGAGGAGCAGCTGCAAGAACATGTTGTTTACCGAAGAAACCTAC 924
 |||||
 QY 721 actcacaacactgaaagcagagatgaatgagttgaaagcgtctccatctcgaagaa 780
 |||||
 Db 925 ACTCCAAACCTGAAACCGACGAGATGATAGTGTGAACAGGCTCCCATTTCCGAGAA 984
 |||||
 QY 781 aacatgttgcctcacaacagaggtgtagagaccacacacacacacacacacacac 840
 |||||
 Db 985 AACCATGTTGGCTCAACCGAGGATGATGAGACCCACCAAGCCCAAGAAACCTCTTCG 1044
 |||||
 QY 841 gtcaactacatgacccaagtcgtcagatgtgacacccaagatgaaagcaggtgcaaggg 900

Db	1045	GTCAACTCATGACCCAAAGTCGTTATGTGATGACCAAAATGAAAGACGGGCAAAAGG	1104
Qy	901	ttcaagcgttaacaggtaccagttcccaagcagctgtccttaaccagaagcgagaattctt	960
Db	1105	TTCCAGGTGTAACAGGTACCGAGTCCAGCCAGCGGCTGCTGACACACAAAGGCAAGATCTTT	1164
Qy	961	ggaagtcgtttctctatgaaagctcgtttaacatatgacgagccgcacatacctaaggaatc	1020
Db	1165	GGAACTCGTTCTATGAAACCTGCTGTAGATATGCGCGCCGCATCACTACAGGGATTC	1224
Qy	1021	cttgaatgaacaagggagagccttggttggatatacaccaaggaacgggaaggtccctcttctgt	1080
Db	1225	CTGATATACCAAGGAGGAGCGCTGGTGATATCACACGGAACGGGAAGTCCCTTCTTGCTG	1284
Qy	1081	aagcgttagaggaacagcgctggcgagtcctccctcagcaaatatacaaccttcagctcaatg	1140
Db	1285	AAGTCTATAGAGACACAGCGCGTGGCAAGTCCCTCACCAAAATCAAAACCTTCAGGCTCATTTATG	1344
Qy	1141	gtgtcaaaagtgaagttgacagatltgagactcttaacagacgtttgctcaactgttgccg	1200
Db	1345	GTCGTCAAAAGTGAAGAGTCGAGATTGTGGACATGCTACAGACCGTTGCTCAGCTGTGCCCCG	1404
Qy	1201	tttgaagaagccagcaactactctgcccagaatccaattgtgccggacacatgcaaaagcga	1260
Db	1405	TTTGAAAGGCGACCAACTACTGCCCCAAGAAATCCATTGTCCGGCACACTGGAAAGACGA	1464
Qy	1261	ccttcctactcagggctcgggtgttttgaagcaaacatctatgtagttactactaagcatctgc	1320
Db	1465	CGTTCTACTAGGGGCTCCGGTGTTTGAAACCAACATCTATGAGAACTTAACCATCTGTCG	1524
Qy	1321	aagacagcgtgtgcaagcgggaatctcaagcaacagagatltgggggttgagcttgagacgtgatg	1380
Db	1525	AAGACAGCGCGTGCACCGCGGAGTCATCAGCAACGAGAGTGGGGGTGACGTGACGTGATG	1584
Qy	1381	ccccgtgataaaaaaagaagaccatgctggagctccg	1413
Db	1585	CCCCGTGATAAAAAGAAAGACTTACACCTGTGCCG	1617

RESULT 12	LOCUS	AX235367	2305 bp	DNA	linear	PAT 11-SEP-2001
DEFINITION	AX235367	Sequence 14 from Patent WO0162928.				
ACCESSION	AX235367					
VERSION	AX235367.1	GI:15593899				
KEYWORDS	.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
AUTHORS	Verret,C.A., Fernandes,E., Shinkets,R.A., Macdougall,J. and Spaderna,S.K.					
TITLE	Polypeptides and nucleic acids encoding same					
JOURNAL	Patent: WO 0162928-A 14 30-AUG-2001;					
FEATURES	Curagen Corporation (US)					
source	Location/Qualifiers					
	1..2305					

BASE COUNT	542 a	619 c	647 g	497 t	ORIGIN
Query Match	86.3%; Score 1286.2; DB 6; Length 2305;				
Best Local Similarity	93.4%; Pred. No 4.5e-297;				
Matches 1393; Conservative	0; Mismatches 3; Indels 95; Gaps				
QY	1 atgagctcgctccctgggtggtgtgtcatccccccttgggctgtctgtcttccgtgtcgtcgtatcc	60			
DB	206 ATGAGCTCTCGTCCGTGGGGGTGTATCCCTTGGGGCGTGTCTTCCTCGGTCCGGGATCC	265			
QY	61 caaaggtaccccccctggcccaacgtctactcccttttagagagtgcttgcgaatatcaacac	120			
DB	266 CAAGGCTATCCTCCGCCCAACGTACCTCTTTAGAGAGGTGCTCCGCAATAACCAAGCAC	325			
QY	121 aacgaftctactcccggtgtccgagaagccatccccaaggagagcaagaggatctctc	180			
DB	326 AAGCAGTCTCACTCCCGGGTCCGAGAGCATCCCAAGGAGGACAAAGAGGAGATCTTC	385			
QY	181 atgtctgcaacaagaagcttcggggcgaaggtgtcacctcccaagctccaaatgtagtaatg	240			
DB	386 ATGTGTGCAACAAGCTTCCGGGCGCAGGTCCACCTCCAGGCTCCCAATGAGATCAATG	445			
QY	241 acctggagatgaagaactgtgagaagctgtctgtcagcgctggcgcaatgacatctcggag	300			
DB	446 ACCTGGAGTACGA-----	459			
QY	301 cagcggtcccaacgaagctgtgtgtgtccatcgtggcagaacctgtggcgctcaactgtggcag	360			
DB	460 -----	470			
QY	361 tatcgctctccgggtgtccatgtgtcagctctgtatgtagcaggtgaagactaacctac	420			
DB	471 TATCGTCTCTCCGGGTTTCATGTGCACTGTCTGTATGACAGAGTAAAGACTACACCTAC	530			
QY	421 cccctaccgagcgaatgtgcaacccctgtgtgtccagaaggtgtctcgggtccataitgtcac	480			
DB	531 CCTACCCCGAGCGAGTCAACCCCTGTGTCCAGAGAGTGTCTCGGGCCATATGTCACG	590			
QY	481 caactacacagatgaatttgggcacccaacaagaatcggtttggctgttaaacactgc	540			
DB	591 CACTACACACAGATAGTTTGGGCCAACACCAACAAAGATCGGTGTGTGTATGCAACCTGC	650			
QY	541 cggaaatatcgtctcgtgggagaagtttgggagaaacggctcacttgtctcgaattat	600			
DB	651 CGAAATATACGTCTGTGGGGAGAAATTTGGGGAACGCGGTCTACTTGTTCGAATAT	710			
QY	601 tctccaaagggaactgtgattggaagaaccccccaagaatgtgcggccctgtctgag	660			
DB	711 TCTCCAAAGGGAACGTGATTGGAGAACCCCTCTTCAAGAAATGGCCGCTCTGTGAG	770			
QY	661 tggccaccagctatgtgaggtcagctgtcagaacaactgtgttacccgagagaaacctac	720			
DB	771 TGCACACCCAGTAAAGGAGGCGAGCTGCAGGAACAATTGTATCCGAGAAAACTTAC	830			
QY	721 actccaaaacccgaaagcagaatgaatgaatgtgaaaacggtcccatctctgaagaa	780			
DB	831 ACTCCAAAACCTGGAAGCGAGCATGATGAATGAGTGGGAACGGGTCCCATTCCTGAAGAA	890			
QY	781 aaccaatgttggcttccaaacgaaggtgtatgtagaaccaacaagcccaagaanaacctctgc	840			
DB	891 AACCAATGTTTGGCTTCCAAACCGAGGATGATGAGACCCCAAGGCCCAAGAAACCTCTGG	950			
QY	841 gtcacactcatgagaccgaagtgtagtgtgtgacacccaagtgtgaagacaggtgtcaaggg	900			
DB	951 GTCAACTACATGACCCCAAGTGTGATGATGTGACCCCAAGATGAAGGACAGGTCCAAGGG	1010			
QY	901 tccacgtgttaacaggtaccaggtgccacagcagctgtccttgaacacaagcggaagatcttt	960			
DB	1011 TCCACGTGTATACAGTACAGTACAGTGCCTGAGTGCCTGTGACCAACCAAGCGGAAGATCTTT	1070			
QY	961 ggaagctgttctctatgaagcgtgtcagcatatgcccgcgcgcacatccactaaggtatc	1020			


```

Db 1071 GGAACCTGTTCTATGAAAGCTCTTACGATATGCGCGCCGCCATTCACATCGGATC 1130
Qy 1021 ctgagatcaaggaggagcctgttgatataccaggagcggagagcccttctgtg 1080
Db 1131 CTGATGACAAAGGAGGAGCTGTGTGATATCCAGGAACGGAGAGTCCCTTCTTGTG 1190
Qy 1081 aagctctagagacagagcgtgagtcctccaggacaatacaacccttccagctcatg 1140
Db 1191 AAGTCTAGAGACACGGCGTGACAGTCCCTGAGCAATATCAACCTTCAGCTATCATG 1250
Qy 1141 gtgcacaagtgaagtgcagagatttgactgtctacacgacccgttgcacgtgtgcg 1200
Db 1251 GTGCAAAAGTGAAGTGCAGAGATTGTGACTACACGACCGTGTCTCACTGTGCCG 1310
Qy 1201 ttgaaaaagcgaacacacacacacacacacacacacacacacacacacacacac 1260
Db 1311 TTTGAAAAAGCAGCACTACTACTGCCCAAGATCATTTGTCGGCAGACACTCAAGAG 1370
Qy 1261 ccttcctactgagctccggtgtgttggaaccaacacacacacacacacacacac 1320
Db 1371 CTTCTACTAGGCTCCGCTCGGTGTGGAACCAACATCTATCATCTCAACATCTGC 1430
Qy 1321 aagacagctgtgacgagagatcatcaagacagagagtggtgagcgtgagcgtgatg 1380
Db 1431 AAGACAGCGGTGCACGCGGAGATCATCAGCAAGAGTGGGCTGACGTGACGTGATG 1490
Qy 1381 cccgtgataaaagaagacacacacacacacacacacacacacacacacacacac 1440
Db 1491 CCGGTGATTAAGAAAGACCTACGTGGGCTCCCTCAGGAATGAGTTCAGTGAAG 1550
Qy 1441 ctggagacacacacacacacacacacacacacacacacacacacacacacacac 1491
Db 1551 CTGGGACCTCTCGGATGGAAGGCGCTTCGGGATCTTGTGTGTCAGGAG 1601

RESULT 13
AK027395 1690 bp mRNA linear PRI 15-MAY-2001
LOCUS Homo sapiens CDNA FLJ14489 f1s, clone MAMMA1002881, weakly similar
DEFINITION to GLIOMA PATHOGENESIS-RELATED PROTEIN.
ACCESSION AK027395
VERSION AK027395.1 GI:14042039
KEYWORDS oligo capping; f1s (full insert sequence).
SOURCE Homo sapiens mammary gland CDNA to mRNA, clone_11b:MAMMA1
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (sites)
Isogai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y.,
Nishikawa, T., Nagai, K., Sugano, S., Aotsuka, S., Yoshikawa, Y.,
Matsumura, H., Ishii, S., Kawai, Y., Saito, K., Yamamoto, J.,
Wakamatsu, A., Nakamura, Y., Nagahara, K., Masuho, Y. and Sasaki, N.
TITLE NEDD human cDNA sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1690)
AUTHORS Isogai, T. and Otsuki, T.
JOURNAL Direct Submission
TITLE Submitted (10-MAY-2001) Takao Isogai, Helix Research Institute,
COMMENT Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3951, Fax: 81-438-52-3952)
NEDD human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan: CDNA full insert sequencing;
5'- & 3'-end one pass sequencing and clone selection; Helix
Research Institute (supported by Japan Key Technology Center etc.)
and Department of Virology, Institute of Medical Science,
University of Tokyo.
FEATURES
Source location/Qualifiers
1. 1690
/organism="Homo sapiens"
/db_xref="taxon:9606"

```

```

CDS
/clone="MAMMA1002881"
/tissue_type="mammary gland"
/clone_11b="MAMMA1"
/feature="cloning vector: pME18FLJ3"
192..1307
/feature="unnamed protein product"
/codon_start=1
/protein_id="BAB55081.1"
/db_xref="GI:14042040"
/translation="MSCVIGGVTPGLGLFLVCGSGYLLPNVTLLELLSKYOHNEH
SVRRRAIPREKEELIAMLHNRKGVQVQASNMEDDLESAAMASQCIWEIG
PTSLIVSIGONIGAMRGVRSRSPGVQVSDVDYRYPSPBCNPPRECSPPKCT
HYTOIVMAFTNKGAVNTRCKRMVGSWENAVFYCNISPKGMVIGEARKNRRC
SECPSPYSGSRRNHCYREYTYTPPEPDEENVEVATIPENHVMLOPRMPTPK
KTSANVNTQVVRCDTKMKDKSCNRYCCPAGCCLNHRKAKIGTIFYESSISICRA
AHHGILDDKGLVDITNGKVPFVKSESRHGVSLR"
BASE COUNT 410 a 469 c 489 g 322 t
ORIGIN

Query Match 74.4% Score 1109.6; DB 9; Length 1690;
Best Local Similarity 99.6%; Pred. No. 7.8e-255;
Matches 1112; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 atgagctgctcctggtgtgtgtatcccttgggagctgtcttctgtctgtgagatcc 60
Db 192 ATGAGCTGGCTCTCTGGTGTGTATCCCTTGGGAGCTGTCTTCTGTCTGCGGATCC 251
Qy 61 caagctacctctgtcccaagtcacatctctagaagagctgtctcaagaataacagac 120
Db 252 CAAGCTACTCTCTCTCCCAACCTCATCTTATAGAGAGCTGTCTCAGCAAAATACAGCAC 311
Qy 121 aacgagctcactcccggttcgagagacatcccaagagagagagagagagatctc 180
Db 312 AACGAGTCTCACTCCCGGCTCCGAGACCATCCCAAGAGAGAACAGAGATCTCTC 371
Qy 181 atctgtcaacaagaagcttcgaggtgagagctcagagcctccacaatagatgacatg 240
Db 372 ATGCTGCAACAAAGCTTCGGGGCCAGGTGCAGCTCAGGCTCCCAACATGATGATCATG 431
Qy 241 acctgagtgacgaactgtggaagatctgtcagcgtgtgagcagatgacatgagag 300
Db 432 ACGTGGATGACGAAGCTGGAAGAGCTGCTCAGCTGGGCGCAAGTCAATGATCTGGAG 491
Qy 301 caagggccacacagctgtgtgtcatctcgagagagagagagagagagagagagag 360
Db 492 CAGGGCCACACAGCTGCTGTGTCTGCTGCGGCAAGACTGGGCGCTCACTGGGCGAG 551
Qy 361 tatgctctcggaggttcacatgtgagagctgtgatacagaggtgaagagactacac 420
Db 552 TATGCTCTCCGGGGTTCATGTGCAGTCTGATATGACGAGGTGAAGAGACTACACTAC 611
Qy 421 ccttaccgagcagagtgaaacccctgtgttcagaagagagagagagagagagagag 480
Db 612 CCGTACCGGACGAGTGCACCCCTGGGTCTCAAGAGAGTCTCAGGCGCTATCTGACG 671
Qy 481 cactacacacagatagtttggtccacacacacacacacacacacacacacacacac 540
Db 672 CACTACACACAGATAGTTTGGGCGACACACACAAAGTCGTTGTCTGTGMAACCTGC 731
Qy 541 cggagagatgactgtctgtggaagagatttgagagagagcggctactactgtctgaat 600
Db 732 CGGAAGAGTGTGTGGGAGAGATTGGAGAACCGGCTCTACTTGTCTGCAATTAT 791
Qy 601 tctccaagggagactgtatgtgagagagcccccataagaatgagcggccctgtctgag 660
Db 792 TCTCCAAAGGGAGCTGTGATGGAGAACCCCTTACAGAGATGGCCGCGCTCTCTGAG 851
Qy 661 tgcacccacagctatgagagcagctgagagacacactgtgttaccgagagagacac 720
Db 852 TGCCACCCACAGCTATGAGAGCAGCTGCAGAACAACTTGTGTTACGAGAGAAACTTAC 911
Qy 721 actccaagaactgaagaagagagatgaaatgaggttggaagagcgtcccatctctgaaga 780

```

|||||
 Db 912 ACTCAAAACCTGAAACGGACGATCAAGAGTGGAAAGGCTCCATCTCTGAGAA 971
 Qy 781 aaccatgttggctcccaaccgaggtgatagagcccaaccgaagaacccctctg 840
 Db 972 AACCATGTTGGCTCCACCGAGGTGATGACGCCACCAAGCCCAAGAAACCTCTGGG 1031
 Qy 841 gtcaactacatgaccccaagtcgtcagatgtgacccaagaatgaaggaaggtgcaagg 900
 Db 1032 GTCAAACTACATGACCCAAAGTCGTACATGTGACACCAAGATGAAAGACAGTCAAAAGG 1091
 Qy 901 tcaacgtgtaacaggtacccaggtcccaagaggtctgctgtaacccaagaaggaatctt 960
 Db 1092 TCCACGCTTACAGGTACAGTCCAGTCCAGAGCTGCTGACCAACAGGGAAGATCTT 1151
 Qy 961 ggaagctgtctatagaagctcgtctagataltgacgcgcgcacacacacacacacac 1020
 Db 1152 GGAACCTGTTCTATGAAACCTGCTAGCATATGCGCGCCGCCCATCTACGGGATC 1211
 Qy 1021 ctgagatgacagagagagctcgtggtatataccaggaaggaaggtccctctctgtg 1080
 Db 1212 CTGGATACCAAGGAGGCTGTGTGATATCACCAGGAAGGAGTCCCTCTCTG 1271
 Qy 1081 aagctgagagacagcagcgtgacgtccctcagcaaa 1116
 Db 1272 AAGTCTGAGAGACACGGCGCTGCACTCCCTCAGTAA 1307

RESULT 14
 AF109674 3052 bp mRNA linear ROD 27-SEP-2000
 LOCUS complete cds.

DEFINITION Rattus norvegicus late gestation lung protein 1 (Lg11) mRNA,
 complete cds.
 ACCESSION AF109674
 VERSION AF109674.1 GI:4324681
 KEYWORDS
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Granulata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE 1 (bases 1 to 3052)
 AUTHORS Kaplan, F., Ledoux, P., Kassamali, F.O., Gagnon, S., Post, M.,
 Koehler, D., Delmaling, J., and Swezey, N.B.
 TITLE A novel developmentally regulated gene in lung mesenchyme: homology
 to a tumor-derived trypsin inhibitor
 JOURNAL Am. J. Physiol. 276 (6), L1027-L1036 (1999)
 MEDLINE 99292450
 PUBMED 10362728

REFERENCE 2 (bases 1 to 3052)
 AUTHORS Kaplan, F. and Swezey, N.B.
 TITLE Direct Submission
 JOURNAL Submitted (26-NOV-1998) Human Genetics, McGill University- Montreal
 Children's Hospital Research Institute, 2300 Tupper Street,
 Montreal, QC H3H 1P3, Canada

FEATURES
 source
 1..3052
 location/Qualifiers
 /organism="Rattus norvegicus"
 /strain="Wistar"
 /db_xref="taxon:10116"
 /cell_type="fibroblast"
 /tissue_type="lung"
 /dev_stage="20 day old fetus"
 1..3052
 /gene="Lg11"
 /note="hormonally and developmentally regulated"
 265..831
 /gene="Lg11"
 /note="Expressed in late gestation lung mesenchymal cells
 in rat, mouse, and human; similar to tumor-derived trypsin
 inhibitor"
 /codon_start=1
 /product="late gestation lung protein 1"

CDS

|||||
 Db 912 ACTCAAAACCTGAAACGGACGATCAAGAGTGGAAAGGCTCCATCTCTGAGAA 971
 Qy 781 aaccatgttggctcccaaccgaggtgatagagcccaaccgaagaacccctctg 840
 Db 972 AACCATGTTGGCTCCACCGAGGTGATGACGCCACCAAGCCCAAGAAACCTCTGGG 1031
 Qy 841 gtcaactacatgaccccaagtcgtcagatgtgacccaagaatgaaggaaggtgcaagg 900
 Db 1032 GTCAAACTACATGACCCAAAGTCGTACATGTGACACCAAGATGAAAGACAGTCAAAAGG 1091
 Qy 901 tcaacgtgtaacaggtacccaggtcccaagaggtctgctgtaacccaagaaggaatctt 960
 Db 1092 TCCACGCTTACAGGTACAGTCCAGTCCAGAGCTGCTGACCAACAGGGAAGATCTT 1151
 Qy 961 ggaagctgtctatagaagctcgtctagataltgacgcgcgcacacacacacacacac 1020
 Db 1152 GGAACCTGTTCTATGAAACCTGCTAGCATATGCGCGCCGCCCATCTACGGGATC 1211
 Qy 1021 ctgagatgacagagagagctcgtggtatataccaggaaggaaggtccctctctgtg 1080
 Db 1212 CTGGATACCAAGGAGGCTGTGTGATATCACCAGGAAGGAGTCCCTCTCTG 1271
 Qy 1081 aagctgagagacagcagcgtgacgtccctcagcaaa 1116
 Db 1272 AAGTCTGAGAGACACGGCGCTGCACTCCCTCAGTAA 1307

/protein_id="AA016986.1"
 /db_xref="GI:4324682"
 /translation="MLINKLRGOVYPPASNMETWDELEERSAAMROCLMEHGA
 SLVLSIGONLAHEWGRTRSPFHYOSYDEKTYITPYPIECNPKPERSGAMCHY
 TOMYMATNTKIGCAVHTRSMVWGDIEVAAYLVNYSFKGMWIDAPYKHPDSE
 CPSSYGGGCGNNLCYREHYHOKPEPMR"
 BASE COUNT 695 a 815 c 851 g 691 t
 ORIGIN
 Query Match 65.3%; Score 974.2; DB 10; Length 3052;
 Best Local Similarity 79.2%; Pred. No. 1.9e-22;
 Matches 1181; Conservative 0; Mismatches 308; Indels 2; Gaps 2;
 1 atgagctgctcctggtggtggtgacacccctggggcctgctcctgctgctgagctc 60
 Db 86 ATGAGCTGCTCTGTGAACATATGTTCCCGGTGGGGCTGGCTCCCTGTGTGCGAGTA 145
 Qy 61 caagctacccctcctcccaacgtcactcctctagaagagctgctcagaataaccagac 120
 Db 146 CAGGCTTTTCTCTCCCAACACATGAGCCTGAGAGGCTGCTCAGCAAGTACAGCAC 205
 Qy 121 aacagatcactcccggttcgcagagacatcccaagaggaaggaagagagatctc 180
 Db 206 ACAGACCAACACTCACGCTGTGGCGAGGCGCATCCCATGTACAGC-CGTCAAAATCTTC 264
 Qy 181 atgctgacacaagaagcttgcgggccaaggtgacgtcagcctcagcctccaacatgagatag 240
 Db 265 ATGCTGACACAAAGTGGAGGTGATGATGCCCTGCTCCCAACATGAGTACATG 324
 Qy 241 acctggatgacgaactgtagaagctcgtcagcgtggtggtggtggtggtggtggtg 300
 Db 325 ACCTGGATGAGAGAGCTGAGAGGTGCGAGCGGCTGGGCGACAGAGGCTGTGGAG 384
 Qy 301 caaggcccccacagctgctggtggtggtggtggtggtggtggtggtggtggtggtg 360
 Db 385 CATGGGCGCGCCAGCCTTCTGTCATCGGGCAGAACCTGCTGTGACCTGGGCGAG 444
 Qy 361 tctgctctccgggtggtcagatgacgtcgtggtggtggtggtggtggtggtggtggtg 420
 Db 445 TACCGCTCTCTGGTTCACAGTGCAGTGTGTATGAGAGTGAAGATTACCTAC 504
 Qy 421 ccttaccgagagaggtgacacccctggtgtcagagaaggtggtcgggtggtggtggtg 480
 Db 505 CCGTACCCCGCAGAGTGAACCCATGATGCTGCCGAGCGTCTCAGAGACCATGTGAC 564
 Qy 481 cactacacacagatggttgggtggtggtggtggtggtggtggtggtggtggtggtg 540
 Db 565 CACTACACACAGATGTTGGTGGCCACACCAACCAAGATTGGCTGTGACACACTGT 624
 Qy 541 cggagaatgactgctggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 600
 Db 625 CGGAGCATGTGAGCTGTGGGAGACATCTGGGAACCGGTGATCTGTGCAATAT 684
 Qy 601 tctccaaaggagagctggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 660
 Db 685 TCTCCAAAGGGAACCTGATGTTGGAGGCTCCTTACAAACATGCGCGCCCTGCTGAG 744
 Qy 661 tgcacacacacatggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 720
 Db 745 TGTGCTGTCAGATGTGAGGTGCTGCGGAGAACACTGTTTACGGAAGAGCACTAC 804
 Qy 721 actccaaacctgaagagcagatgaaatgaaggtggtggtggtggtggtggtggtggtg 780
 Db 805 CATGAGAACCC-GGAGTGTGATGATGAACGAGGTGGAATCACCCCTGCTCCGAGGA 863
 Qy 781 aaccatgttggctcccaaccgaggtgatagagcccaaccgaagaacccctctg 840
 Db 864 ACCCATGTTGGTTCACCAAGGTGTGTAACACCTCAAAAGCAAGAACCCCAAGTC 923
 Qy 841 gtcaactacatgaccccaagtcgtcagatgtgacccaagaatgaaggaaggtgcaagg 900
 Db 924 GTCAACTTCATGACCCAAAGTGTCTCACTGTGACCAACAAAGTGAAGACTCATGCAAGGA 983


```
OY 1163 atttgactgtacagaccgttctcagctgtgccgtttgaaaagccagcaactcact 1222
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1284 CTGTGACTTGTGAAACACTGTGTGAAACAGCTCTGTTCATTTCATTAAGCTGTTCACATT 1343
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1223 gcccaagaatccatgtctcgcgcacactgcacaaagcgaaccttcctactggtccggtgt 1282
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1344 GCCCAAGAGTATACGTGCTGTGTAATGCAAGCAAAATCCACATTATGCTGTGTAA 1403
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1283 ttggaaccaacatctatgcagatacctcaagcatctgcagaagcgtgtgcacgcggag 1342
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1404 TTGGAACTCGAGTTTATCTGTATCTGCAGTATCTGCAGAGCAGACATATCTGTGAG 1463
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1343 tcatcaagaacgagagtgagggtgacgtgcgtgcgtgcgtgcgtgcgtgcgtgcgtgcgt 1402
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1464 TGGTTCGAAATCA---CGGTGTATATGTGATGTATATCCCTGTGTGACAAAAGAAAGACT 1520
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1403 acgttggtctgcctcagaatgagttcagtcctgaagccttgaggactcctcggatgaa 1462
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1521 ACATTGCTTCTTTCAGAAATGGAATCTCTCAGAAAGTTTACAGATATCTCTCAGAGGAA 1580
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1463 aggccttcggatcttgcgtgt 1484
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1581 AGGCATTTCAGAGTGTGTCTGT 1602
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: May 25, 2002, 13:22:02
Job time: 10318 sec

Tue May 28 16:10:02 2002

us-09-667-380a-1.rst

No suitable data
Page 1

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2002, 10:25:14 ; Search time 2064.9 Seconds
(without alignments)
9745.731 Million cell updates/sec

Title: US-09-667-380A-1
Perfect score: 1491
Sequence: 1 atgagctgcgtctgctggtg.....ggatcttcgtctcagcag 1491

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 674847542 residues
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: em_estba:*
2: em_estbun:*
3: em_estlin:*
4: em_estlun:*
5: em_estlov:*
6: em_estlpl:*
7: em_estro:*
8: em_estt:*
9: gb_estl:*
10: gb_estl2:*
11: gb_hlc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_hiv:*
15: em_gss_pln:*
16: em_gss_vrl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	967	64.9	2929	AK019034	Mus muscu
2	509	34.1	651	BF361052	MR2-OT007
3	447.2	30.0	482	BG011536	PM1-GN018
4	439.8	29.5	459	BF998318	OV0-GN021
5	434.2	29.1	900	BG175425	602337760
6	421.4	28.3	680	BI911002	603068720
7	411.8	27.6	748	BI855378	603382445
8	401	26.8	835	BF181843	601805465
9	398.2	26.7	580	AW910974	ur81f08.y
10	398.6	26.7	762	BI155480	602903864
11	395	26.5	406	BI030828	IL5-MT026
12	391.2	26.2	822	BF141923	603315836
13	387.8	26.0	783	BF1889391	OV3-CR055
14	383.6	25.7	763	BG681553	60295324
15	379.2	25.4	791	BI248527	602992844
16	375.8	25.2	791	BI248527	602992844
17	358.8	24.1	573	BE310264	601089361

18	358.6	24.1	857	10	BI687001	BI687001 603313888
19	356.8	23.9	521	9	AI549820	AI549820 we81d05.y
20	344.4	23.1	483	10	BF151925	BF151925 uc22a09.y
21	336.4	22.6	738	10	BF026829	BF026829 601670724
22	323.2	21.7	693	10	BF662117	BF662117 603305133
23	323.2	21.6	360	10	BF355122	BF355122 RC6-HT083
24	321.6	21.5	980	10	BF161588	BF161588 601771087
25	302.4	20.3	937	10	BE308689	BE308689 601097250
26	302.2	20.3	768	10	BF166647	BF166647 601776955
27	299.6	20.1	447	9	AA423729	AA423729 ve81d05.y
28	298.4	20.0	582	10	BI697147	BI697147 603346540
29	297	19.9	1040	10	BI855520	BI855520 60326145
30	296	19.9	664	10	BI855155	BI855155 603382183
31	280.2	19.5	515	10	BI112026	BI112026 602901366
32	285.4	19.1	650	10	BE572064	BE572064 601333817
33	283.4	18.5	598	9	AM321284	AM321284 uc29e12.y
34	276.2	18.4	456	10	BF375908	BF375908 PM4-TN002
35	274.6	18.2	668	10	BI663711	BI663711 603288431
36	271	17.5	557	10	BF310707	BF310707 601089662
37	260.2	17.2	370	10	BF754227	BF754227 IL5-CR051
38	259.8	17.0	358	10	BF772739	BF772739 IL5-TT002
39	256	16.9	289	10	BF854672	BF854672 MR2-EN009
40	253.6	16.9	352	10	BI105487	BI105487 602891816
41	252.2	16.2	824	10	BG920333	BG920333 602825837
42	241	16.0	343	10	BE532699	BE532699 601234335
43	238.2	15.6	801	10	BG193229	BG193229 RSM12356
44	233.2	15.6	801	10	BG193229	BG193229 RSM12356
45	231.2	15.5	515	10	BE481031	BE481031 166323 BA

ALIGNMENTS

RESULT 1	Accession	Version	Keywords	Source	Organism	Reference	Authors	Title	Journal
AK019034	AK019034.1	GI:12859037	HTC, CAP trapper.	Mus musculus (strain: C57BL/6J) 10 day old male pancreas cDNA to enriched library, clone: 1810049K24; homolog to LATE GERSTATION LUNG PROTEIN 1, full insert sequence.	Mus musculus	Carninci, P. and Hayashizaki, Y.	Shibata, K., Itoh, M., Kono, H., Akizawa, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Kashiwagi, K., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Ohara, E., Watanabe, Y., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawak, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, H. and Hayashizaki, Y.	High-efficiency full-length cDNA cloning by RT-PCR	Genome Res. 10 (11), 1757-1774 (2000)
AK019034	AK019034.1	GI:12859037	HTC, CAP trapper.	Mus musculus (strain: C57BL/6J) 10 day old male pancreas cDNA to enriched library, clone: 1810049K24; homolog to LATE GERSTATION LUNG PROTEIN 1, full insert sequence.	Mus musculus	Carninci, P. and Hayashizaki, Y.	Shibata, K., Itoh, M., Kono, H., Akizawa, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Kashiwagi, K., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Ohara, E., Watanabe, Y., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawak, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, H. and Hayashizaki, Y.	High-efficiency full-length cDNA cloning by RT-PCR	Genome Res. 10 (11), 1757-1774 (2000)
AK019034	AK019034.1	GI:12859037	HTC, CAP trapper.	Mus musculus (strain: C57BL/6J) 10 day old male pancreas cDNA to enriched library, clone: 1810049K24; homolog to LATE GERSTATION LUNG PROTEIN 1, full insert sequence.	Mus musculus	Carninci, P. and Hayashizaki, Y.	Shibata, K., Itoh, M., Kono, H., Akizawa, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Kashiwagi, K., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Ohara, E., Watanabe, Y., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawak, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, H. and Hayashizaki, Y.	High-efficiency full-length cDNA cloning by RT-PCR	Genome Res. 10 (11), 1757-1774 (2000)


```

QY 1021 ctgataagcaagagagcctgttgatatacccaagcaagagagctccctctctg 1080
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 1217 ATCGATGATCGAGTGCGCTGGTGTGATGTCACACAGGAATGGGATGGTCCCTTCTTC 1276
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
QY 1081 aagctgagagacacagcgtgagtcctcctcagcaatacaactcctcagctcattc 1140
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 1277 AAGTCTCAGAAAAATGCGATGAGTCCCTGACCAATACMACCCGTCTGACTCTTCACT 1336
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
QY 1141 ggtcaaaagtgaagtgtgagagattgagctctacagcagcgtgtgctgctgccc 1200
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 1337 GGTCTCAAAAGTACACAGAGCGCCGTGGATGCTGACACGCCAGCTGACAGTGTGCCCC 1396
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
QY 1201 ttgtaaaagccagcaactacgtcccaagaalccattgtccggcacactgcaaaagca 1260
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 1397 TTGGAAGACCGGCCACCCACCTGCGGAGAAATCCAGTGTCTGCGCGGTGTGAGAGAG 1456
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
QY 1261 ccttctactgagctcgtgtgttttgaaaccaatctatgacagatacctcagacatctgc 1320
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 1457 CCATCTATTGGGCTCTCTGTATGAGACCCACATCTATGCTGATCTTCCAGCATTTGT 1516
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
QY 1321 aagagcgtgacagcagcagagtcacagcaagagagtgagtgagcgtgagtgatg 1380
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 1517 AAGGCGCTGTGACGCCAGAGTGTCTGACGAGGTTGCTGATGAGATGTGATG 1576
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
QY 1381 cccgtgataaaaaaagaagacctcgtgggctgctcaggaatgagtgagtgagtgag 1440
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 1577 CCCGTGACAAAAAAGAGACCTACGTGGCTCCCTCAGGAAGGAGGAGGAGTGGAGAGC 1636
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
QY 1441 ctggagagctctcaggaatggaagagcctcagagctctgctgctgagcag 1491
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 1637 CTGACACTCTCAGAACGGAAGCGCTTCCGGATCTTCCCGCTGAGAGCAG 1687
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||

```

```

RESULT 2
BF361052/c 651 bp mRNA linear EST 22-NOV-2000
LOCUS BF361052
DEFINITION M82-OT0079-230500-009-a08 OT0079 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF361052
VERSION BF361052.1 GI:11320124
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 651)
AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?pl=MR2&t2=MR2-OT0079-
230500-009-a08&t3=2000-05-23&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 561.
Location/Qualifiers
1..651
/organism="Homo sapiens"

```

```

FEATURES
source
1..651
/organism="Homo sapiens"

```

```

/db_xref="taxon:9606"
/clone_lib="OT0079"
/dev_stage="Adult"
/note="Organ: ovary; Vector: puc18; Site: 1; SmaI; Site-2:
SmaI; A mini-library was made by cloning products derived
from ORSTES PCR (U.S. Letters Patent application No. 196
'716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
BASE COUNT 150 a 179 c 161 g 161 t
ORIGIN

```

```

Query Match 34.1%; Score 509; DB 10; Length 651;
Best Local Similarity 96.9%; Pred. No. 3e-119;
Matches 540; Conservative 0; Mismatches 15; Indels 2; Gaps 2;

```

```

QY 932 gctgctgaacccaagcgaagatccttggaatctgttctatgaaagctcgtcagca 991
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 639 GCTGGCTGAAACACAAAGCGAAGATC-TTGGAACTCTGTCTATGAAAGCTCGTCTAACA 581
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
QY 992 -tatgcgcgcgcgcacacacacacacacacacacacacacacacacacacacacac 1050
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 580 TTATTTCGCGCCCAATTCTACACGGATCCTGATGACAAAGGAGGCGCTGATGATTC 521
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
QY 1051 accaggaacggaaggtgcccttctcgtgaaatctgagagacagcggtgacgtccctc 1110
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 520 ACCAGGAAGGGAAGGATCCCTCTCTCGAAGTGTGAGACACAGCGGTGACATCCCTC 461
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
QY 1111 agcaaatataaaccttcacacacacacacacacacacacacacacacacacacacac 1170
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 460 AGCAAAATACAAACCTTCCAGCTCATTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 401
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
QY 1171 tgcctacagacgctgtgctcagctgtgcccgtttgaaagccagcaactacatcgcacaga 1230
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 400 TGGTACACGACCGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 341
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
QY 1231 atccatgttcacgacacacacacacacacacacacacacacacacacacacacacac 1290
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 340 ATCCATTTGCTCGGACACCTGCAAGAGCAACCTTCTACTAGGGCTCCGGGTTGGAACC 281
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
QY 1291 aacatcatgacatacctcaagacatctgcaagacagctgtgacagcgagatcaccagc 1350
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 280 AACATCTATGACAGTACTTCACGATCTGCAAGACAGCTGTCGCCCGGAGATCATCAGC 221
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
QY 1351 aacgagagtgagggtgacgctgacgctgacgctgacgctgacgctgacgctgacgctg 1410
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 220 AACGAGAGTGGGGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 161
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
QY 1411 tcgctcaggaatgagtgacgtcgtgaaagccttgggagctcctcgggagtgaaagccttc 1470
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 160 TCGCTCAGAGATGGAATGCAATCGAAGACCTGCGGAGCTCTCGGATGGAAGGCCCTTC 101
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
QY 1471 cgaatccttgcgtcag 1487
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 100 TGGATCTTGTGCTGTACG 84
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||

```

```

RESULT 3
LOCUS BG011536 482 bp mRNA linear EST 24-JAN-2001
DEFINITION BG011536
ACCESSION BG011536
VERSION BG011536.1 GI:12459833
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 482)
AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,

```


QY 1188 tcagctgtgcccgtttgaaagccagcaactactgcccagaatccatgtgccgacaca 1247
 DB 279 TCAGCTGTGCCCGTTTGAAGCCAGCAACTCAGTCCCAAGAAATCCATTTGTCCGGACACA 220
 QY 1248 ctgcaagaagcaactctactactggtcgcggtgtgttggaaaccaatctatgcagatac 1307
 DB 219 CTGCAAAAGCAAGCTCTCTACTGTGGGCTCCGGTGTGGAAACCAACATCTATGCAATAC 160
 QY 1308 ctcaagcatctgcaagaagcgtgtcaagcggtggtatcatgcaagcagagatgtgggttca 1367
 DB 159 CTCAAGCATCTCAAGACAGCTGTGTACAGCGGGAGTCTATCAGCAACGAGAGCGGGGTGA 100
 QY 1368 cgtgagcgtgagcccggtgatacaaaagagacactactgtgggtcgtcgaagatgtgagt 1427
 DB 99 CGTGACGATGATGCCGTGTGATTAAGAAAGACCTACGTGGGCTGCTCAGGAAATGAGAT 40
 QY 1428 tcagctgaaagcctgaggactcctcggatgtaa 1462
 DB 39 TCAGCTGAAAGCCTGGGG-CTCCTGGGATGGA 6
 RESULT 5 900 bp mRNA linear EST 06-FEB-2001
 BG175425
 LOCUS 602337780F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:4460802 5'
 DEFINITION mRNA sequence.
 BG175425
 ACCESSION BG175425.1 GI:12682128
 VERSION EST.
 KEYWORDS house mouse.
 SOURCE Mus musculus
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 900)
 NIH-MGC http://mgs.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: LLM10262 row: P column: 19
 High quality sequence stop: 686.
 Location/Qualifiers
 1..900
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:4460802"
 /clone_1db="NCI_CGAP_Mam1"
 /tissue_type="tumor, biopsy sample"
 /dev_stage="10 months, virgin"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: PCMV-SPORT6; Site: 1: Saliv;
 Site: 2: Noli; Cloned unidirectionally. Primer: Oligo dT.
 Library constructed by Life Technologies. Investigator
 providing samples: Gilbert Smith, NIH"
 BASE COUNT 175 a 248 c 278 g 199 t
 ORIGIN
 Query Match 29.1%; Score 434.2; DB 10: Length 900;
 Best Local Similarity 80.7%; Pred. No. 4,4e-100;
 Matches 519; Conservative 0; Mismatches 123; Indels 1; Gaps 1;
 QY 849 catgaccagtgctcagatgtgacaccaagatgaagacaggtgcaaggtccacgtg 908
 ||||||||||||| ||||||||||||||| ||||||| |||||||

DB 10 CATGACCCAGTGGTCCACTGTGTACACCAAGATGAGAGACTCATGCAAGATCCAGTGG 69
 QY 909 taacaggtaccagtggtcccaagagctgtgcaacacaaagggcaagatcttggagctc 968
 DB 70 TAACAGGTACCAGTGGTCCACAGAGCTGTGTGACGACACAAAGGAGTCC-TTGGCTCTCT 128
 QY 969 gtctatgaaagctcgtctgcatatgcccgcgcgcgcacacacacagggatcctgatatga 1028
 DB 129 GTTTATGAAGTCTCTTCCAGCATATCCAGAGCTGTATCCACTACAGCTGTATCATGTGA 188
 QY 1029 caagggagcctgtgtgatatcaccaggaacgggaaggtccctcttcttggagcttga 1088
 DB 189 TCAGAGTGGCTGTGTGTGATGTCCACAGAGATGGATGTGTACCTCTTGTGCAAGTCTCA 248
 QY 1089 gagacagcggtgacgtccctcccaagaaatacaaaccttccagctcatcattggtgtcaaa 1148
 DB 249 GAAAATATGACATGAGTCCCTGAGCAATATCAAGCCGCTCTGCTCTCTCTGCTGAAA 308
 QY 1149 agtgaagctgcaagatttgatctgctcaagacggttgcagctgtgcccgttggaaa 1208
 DB 309 AGTGACAGAGACGGCGCTGTGAGTCCACAGCGGCTGACAGCTGTGCTCCCTTCAGAA 368
 QY 1209 gccagcaactcactgcccagaatccattgtccggtgacactgtgcaagaacgaactccta 1268
 DB 369 GCCAGCACCCACTGCTGCCAGAAATCAGATGCTCTGCGGCTGTGAGAAAGCATCTCTA 428
 QY 1269 ctgggtcctcgtgtgttggaaaccaatctatgacagatcctaagacatctgcaagaagc 1328
 DB 429 TTGGGCTCTGTGTATGAGAACCAACATCTATCTTCCACATTTGTGAAGCCGCC 488
 QY 1329 tgtgacgctggagatcatgaagaacagagtgagggtgacgtgacgtgacgtgacgtgac 1388
 DB 489 TGTGACGACGAGTGTATGTTGACAGAGTGTGTGCTATGAGATGTGATGCCCTGGA 548
 QY 1389 taaagaagacactactgtgtgctgctgcaagatgagttcagttcgaagcttggagac 1448
 DB 549 CAAAAGAAAGCTAGTGTGGGCTCTCTCAGAGAAAGGGGCTGAGAGAGCTGTAAC 608
 QY 1449 tccctggatgtaagagcctcctcgatcttctgtgtcagagcag 1491
 DB 609 TCCTCAGAACGAAAGCCCTTCGAGATCTTCCGCTCAGGACAG 651
 RESULT 6 680 bp mRNA linear EST 16-OCT-2001
 BI911002
 LOCUS 603068720F1 NIH_MGC_118 Homo sapiens cDNA clone IMAGE:5217820 5'
 DEFINITION mRNA sequence.
 BI911002
 ACCESSION BI911002.1 GI:16174517
 VERSION EST.
 KEYWORDS human.
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 680)
 NIH-MGC http://mgs.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Life Technologies, Inc.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: LLM1547 row: g column: 05
 High quality sequence stop: 675.
 Location/Qualifiers
 1..680
 /organism="Homo sapiens"
 FEATURES
 source

FEATURES
 Tissue Procurement: Jeffrey Green M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/BLN at:
<http://image.llnl.gov>
 Plate: LAM12058 row: 1 column: 12
 High quality sequence stop: 748.
 Location/Qualifiers

BASE COUNT	164 a	230 c	212 g	142 t
ORIGIN	Samples: Jersey Green, M.D., NIH			

Query Match	27.68;	Score 411.8;	DB 10;	Length 748;
Best Local Similarity	78.5%;	Pred. No. 2.1e-94;		
Matches 505; Conservative	0;	Mismatches 137;	Indels 1;	Gaps 1

Db	406	CACGGGCCCCCGCCCTTGGAGGTGCATACGGGCGAAGCACTGGCTGTGCACTGGGGCAGG	46
QY	361	TATGCTCTCCGGGTTTCATGTGCAAGCTCTGATAGCAAGTGTGAAGACTACACTAC	42
Db	466	TACGGCTCTCTCGTGGTTCATGTGCAATGTGTAACGACAGGGGTGAAGGATTACACTAC	52
QY	421	CGCTAACCGGAGCGAGTGTGAACCCCTGTGTCTCAGAGAGTGTCTCGGGCTATGTGAAG	48
Db	526	CCATATCCCCACGAGTGCACTCAACAGGTGCGGGGAGGGGTCTCAGGGCCCATGTGTGACC	58
QY	481	CACCTAACACAGATAGTTGTGGCCACCAACAAGAATGTTGTGTCTGTGAACCTGTC	54
Db	586	CACCTAACACACAGATGTTGTGGGCCACGACCAACAAATCGCTGTCCGTGCACACCTGT	64
QY	541	CGGAGATGACGTGTGGGGGAGAAATTTGGAGAAAGCGTCTACTCTGTCTGCATAT	60
Db	646	CGGAAATATTAAGCTGTGGGGAGACACTTTGGGAATAATGCCGTATCTCGTGTGCATTTAT	70
QY	601	TCTTCCAAGGGGAAGCTGTGATTTGAGAAAGCCCTCTCAAGAAAT	642

Db 706 TCTCCCAAGGAACTGATTGGCGAGGCCCTACTAAGAT 748

RESULT 8 835 bp mRNA linear EST 31-0CT-2000
 BF181843
 LOCUS 601805465f1 NCI-CGAP_Mam5 Mus musculus cDNA clone IMAGE:4056145 5',
 DEFINITION mRNA sequence.
 ACCESSION BF181843 GI:11059985
 VERSION BF181843.1
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 835)
 REFERENCE NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM9311 row: b column: 18
 High quality sequence stop: 672.
 Location/Qualifiers

FEATURES
 source 1..835
 /organism="Mus musculus"
 /strain="C57/B6"
 /db_xref="taxon:10090"
 /clone_image="4036145"
 /clone_lib="NCI-CGAP_Mam5"
 /tissue_type="tumor, gross tissue"
 /dev_stage="7 months"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: pCMV-Sport6; Site:1; Salt;
 Site:2; Notti. Cloned unidirectionally. Primer: Oligo dT.
 Library constructed by Life Technologies. Investigator
 providing samples: Lothar Hennighausen/Robin Humphreys,
 NIH"

BASE COUNT 178 a 237 c 244 g 176 t
 ORIGIN

Query Match 26.9%; Score 401; DB 10; Length 835;
 Best local similarity 80.7%; Pred. NO. 1.3e-91;
 Matches 517; Conservative 0; Mismatches 120; Indels 4; Gaps 4;

QY 842 tcaacatgacccagcagtcgacatgtgacacacagatgagacagagtcacaaagt 901
 Db 8 TCAACTTCATGACCAAGGTGCTCACTGTGACACCAAGGTGAGCTCATCAAGAT 67
 QY 902 ccaagtcacagagtcacagtcgacagagtcgacacacagagtcacaaagt 961
 Db 68 CCAGGTGTACAGGTACCAAGTGTGCTGAGACCAAGGCGAAGGTC-TTG 126
 QY 962 gaagtcctctacgaagtcgctcag-catacgccgcgcgcacacacagtcac 1020
 Db 127 GCTCTCTTTTATGAAGTCTTCCAGCCATATGCGGAGCTGCTATCCACTACGGTGC 186
 QY 1021 ctgagtcacagagtcgctgctgacacacagagtcacaaagtccctctctcgtg 1080
 Db 187 ATCGATGATCGAGGTGCTGCTGATGTCAACGAAGATGGAGTACCCCTCTTTGTC 246
 QY 1081 aagtcgagacagcgcgctgacccctcagcaataacacacacacacacacacac 1140
 Db 247 AAGTTCAGAAAATGCGATGATGCTCCGTGAGCAATACAGCCCTCTAGCTCTTCACT 306

QY 1141 gtgtcaaaagtgaaagtgacagatttgactgtcacagacgctgtcagctgccc 1200
 Db 307 GTGTCAAAAAGTGTACAGAG-ACGGCTGTGACTGCTCCACCGCAGCTGTGCCCC 365
 QY 1201 ttgtgaaagcgaactcactcactcctcctcctcctcctcctcctcctcctc 1260
 Db 366 TTGTGAGAGCCGGCAC-CACTGCCGAGAAATCCAGTCTCTGCGCTGTGTGAGAG 424
 QY 1261 cctctcactgagctcctcctcctcctcctcctcctcctcctcctcctcctcctc 484
 Db 425 CCATCTATTGGGCTCTCTGTGTGACCAACATCTATGTGATCTTCCAGCATTTGT 1380
 QY 1321 aagacgctgacgc 1440
 Db 485 AAGGCCCTGTGACGAGGTCATCTGTTGACGAGGTTGCTGTGCTATGATGTGATG 544
 QY 1381 cccgtgataaaagaaagcactcctcctcctcctcctcctcctcctcctcctcctc 1440
 Db 545 CCGGTGACAAAAGAGAGCTACGTGGCTCCCTCAGAACGGGGTGCAGTCGAGAGC 604
 QY 1441 ctggggaactcctcgcgagtgaaagcctcctcctcctcctcctcctcctc 1481
 Db 605 CTGACACTCTCTCAGAACGAAAGCCTTCGATCTTCG 645

RESULT 9 580 bp mRNA linear EST 25-MAY-2000-
 AW910974
 LOCUS ur81f08.y1 NCI-CGAP_Mam6 Mus musculus cDNA clone IMAGE:3156711 5',
 DEFINITION similar to TR:Q26422 Q26422 FACTOR C. [1] ;, mRNA sequence.

ACCESSION AW910974
 VERSION AW910974.1 GI:8076219
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Jeffrey Green M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov/html/resources.shtml

MG1:1059467
 Seq primer: -40RP from GIBCO
 High quality sequence stop: 428.
 Location/Qualifiers

FEATURES
 source 1..580
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone_image="3156711"
 /clone_lib="NCI-CGAP_Mam6"
 /sex="female, virgin"
 /tissue_type="infiltrating ductal carcinoma"
 /dev_stage="5 months"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: pCMV-Sport6; Site:1; Salt;
 Site:2; Notti. Cloned unidirectionally. Primer: Oligo dT.
 Library constructed by Life Technologies. Investigator
 providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT 138 a 150 c 163 g 129 t
 ORIGIN

AUTHORS
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, M. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeblut, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.

TITLE
 Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

COMMENT
 Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?l=QV3&l2=QV3-CT0556-041000-370-f05&l3=2000-10-04&l4=1>)
 Seq primer: puc 18 forward
 High quality sequence start: 9
 High quality sequence stop: 410.
 Location/Qualifiers
 1. 410
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_1lb="CT0556"
 /dev_stage="Adult"
 /note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT
 109 a 111 c 109 g 81 t

ORIGIN

Query Match 25.7%; Score 383.6; DB 10; Length 410;
 Best Local Similarity 98.8%; Pred. No. 2.5e-87;
 Matches 397; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 1005 catccactcggagatccctgatgacagaggagcctgtgtgatataccaggaagaa 1064
 DB 9 CATCCACTGAGGATCCTGGATGACAGGAGGCTGTGTGATATCACCCAGGAACGGGAA 68

QY 1065 ggtccctctctcgt-gaagctctgagagacagcgctgcagctccctcagaacaatacaac 1123
 DB 69 GGTCCCTCTCTCGTGAAGCTCTGAGAGACAGCGCTGCAGTCCCTCAGCAATATACAAAC 128

QY 1124 ctccagctcatctatgtgtgtaaaagtgaaagtgcaagatttgagactgtactacagccg 1183
 DB 129 CTCCAGCTCATTCATGATGTGTCAAAAGTGAAAGTGCAAGATTTCACCTGTACACGACCG 188

QY 1184 ttgtcgaagctgtcccgctgtgaaagcagaactcactctgcccagaagatcatgttcgg 1243
 DB 189 TTGCTAGCTGTCCCGTTGAAAGCCAGCAACATCTCCTGCCCAAGATTCATTGTCCGG 248

QY 1244 ccaactgcaagaagacactctcactgagctccggtgtgttgaaacaacaatcatatgag 1303
 DB 249 CACACTGCAAAAGAGAACTTCTACTAGGGCTCCGGTGTGGAACCAACATATATGAG 308

QY 1304 atactcctcaagatctgcaagaagctgtgcaagcggaatcacaagaagagtgagg 1363
 DB 309 ATACTCCTCAAGCATCTGCAAGAGCTGTGACGAGAGATCATCAGCAACGAGATGGGG 368

QY 1364 gtgaagctgacgtgatgacggtgatataaaagaagacttacc 1405
 DB 369 GTGACGTGAGACGTGATGCCCGGTGATCAAAAGAAAGACCTTACG 410

RESULT 15
 BG61553

LOCUS
 BG61553

DEFINITION
 BG61553 763 bp mRNA linear EST 29-MAY-2001
 602795324P2 NIH_CGAP_Mam4 Mus musculus cDNA clone IMAGE:4916597 5', mRNA sequence.

ACCESSION
 BG61553

VERSION
 BG61553.1 GI:14212091

KEYWORDS
 EST.

SOURCE
 house mouse.

ORGANISM
 Mus musculus

REFERENCE
 1 (bases 1 to 763)
 NIH-MGC <http://mgs.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1995).

AUTHORS
 CONTACT: Robert Strausberg, Ph.D.
 Email: cga@bcrfemail.nih.gov

TITLE
 Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth Ph.D.

COMMENT
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LIML at:
<http://image.llnl.gov>
 Plate: LLAM10826 row: p column: 06
 High quality sequence stop: 755.
 Location/Qualifiers
 1. 763
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone_1img="4916597"
 /clone_1lb="NIH_CGAP_Mam4"
 /tissue_type="tumor, gross tissue"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: PCMV-SPORE6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: oligo dT. Average insert 2.5 kb. Library constructed by Life Technologies, catalog # 12018-016. Investigators providing samples: Lothar Hennighausen/Priscilla Furth, NIH Reference for transgenic model: Li et al., Cell Growth and Differentiation 7, 3-11 (1996). Note: this is a NCI_CGAP Library."

BASE COUNT
 165 a 238 c 213 g 147 t

ORIGIN

Query Match 25.4%; Score 379.2; DB 10; Length 763;
 Best Local Similarity 78.1%; Pred. No. 4.5e-86;
 Matches 481; Conservative 0; Mismatches 133; Indels 2; Gaps 2;

QY 1 atgagctgcgtccctgggtgtgtcatcccttggggctgtctctcgtgtcgtatcc 60
 DB 149 ATGAGCTGCTCTTCTGAACAATATGTCTGATGGGGCTGCTGTGCTGTGTGAGTA 208

QY 61 caaggtactcctctgcccgaagctcactctcttagagagagctgtctcagcaataaccacac 120
 DB 209 CAGGCTTTTTCCTTCCCAACACACAGGCTGTGAGAGCTGTGAGCAATATACCAATAT 268

QY 121 aacgaatcactcctccgggtccgcaagacatcccccagagagagcaagaagagatccctc 180
 DB 269 GCAGAGCCACACATCAGCGGTGCGAGGCGCATCCCATGTCAAGACGCCCAAGATATCTC 328

QY 181 atgtgtcacaacaagctctcggtgccaaggtgtgcaagcttcacagctccaacatgtgacatg 240
 DB 329 ATGTCTACACAAACAGCTGCGAGGCGCATGTTATCCCTCTCCCAACATGAAACACATG 388

QY 241 acctggatgacgaactgtgagagctgtgtcgtcagcgtgtggccagctcagtcacatctggag 300
 DB 389 ACTTGGATGAGGAGCTAGAGAGATCTGCAGCAGCGTGGGCACATATAGATGCTGTGGAG 448

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2002, 10:33:09 ; Search time 70.13 Seconds
(without alignments)
5222.295 Million cell updates/sec

Title: US-09-667-380A-1
1491
Sequence: 1 atgagctgcgtcctcgtggtg.....ggatcttctcgtcagcag 1491

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:
1: /cgn2_6/prodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/prodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/prodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/prodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/prodata/2/ina/PCUS.COMB.seq:*
6: /cgn2_6/prodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	74.6	5.0	970	2	US-08-773-368-2
2	74.6	5.0	970	3	US-09-199-887-2
3	43.2	2.9	7218	1	US-08-232-463-14
4	42.2	2.8	11219	1	US-07-642-734C-1
5	42.2	2.8	11219	3	US-08-439-009A-1
6	38.4	2.6	696	4	US-08-998-416-689
7	38.4	2.6	30001	1	US-08-125-468-1
8	38.4	2.6	30001	2	US-08-474-933-1
9	37.4	2.5	4031	1	US-08-159-784-1
10	37	2.5	1211	2	US-08-997-080-40
11	37	2.5	1211	2	US-08-997-362-40
12	37	2.5	1211	2	US-08-873-970-40
13	37	2.5	1211	4	US-09-095-855-40
14	37	2.5	1211	4	US-08-705-347A-40
15	37	2.5	1211	4	US-09-324-542-40
16	36.8	2.5	756	4	US-09-175-014-1
17	36.4	2.4	3468	1	US-07-951-715A-2
18	36.4	2.4	3468	2	US-08-459-448A-2
19	36.4	2.4	3468	3	US-08-459-595A-2
20	36.4	2.4	3468	3	US-08-459-504B-2
21	36.4	2.4	3468	3	US-08-459-444-2
22	36.4	2.4	3468	3	US-09-053-549-3
23	36.4	2.4	3468	4	US-09-547-422-2
24	36	2.4	1212	6	5212296-17
25	35.8	2.4	1410	3	PCT-US95-15428-14
26	35.8	2.4	1817	3	US-08-564-496C-14
27	35.6	2.4	696	4	US-08-998-416-1050

28	35.2	2.4	71989	4	US-09-443-501A-2	Sequence 2, Appl1
29	34.8	2.3	1080	2	US-09-205-922-1	Sequence 1, Appl1
30	34.8	2.3	1947	1	US-07-951-715A-3	Sequence 3, Appl1
31	34.8	2.3	1947	2	US-08-459-448A-3	Sequence 3, Appl1
32	34.8	2.3	1947	3	US-08-459-595A-3	Sequence 3, Appl1
33	34.8	2.3	1947	3	US-08-459-504B-3	Sequence 3, Appl1
34	34.8	2.3	1947	4	US-08-459-444-3	Sequence 3, Appl1
35	34.8	2.3	1947	4	US-09-547-422-3	Sequence 3, Appl1
36	34.8	2.3	3468	1	US-07-951-715A-4	Sequence 4, Appl1
37	34.8	2.3	3468	2	US-08-459-448A-4	Sequence 4, Appl1
38	34.8	2.3	3468	2	US-08-459-595A-4	Sequence 4, Appl1
39	34.8	2.3	3468	3	US-08-459-504B-4	Sequence 4, Appl1
40	34.8	2.3	3468	3	US-08-459-444-4	Sequence 4, Appl1
41	34.8	2.3	3468	3	US-08-459-504B-8	Sequence 4, Appl1
42	34.8	2.3	3468	3	US-08-459-444-4	Sequence 4, Appl1
43	34.8	2.3	3468	3	US-08-459-504B-8	Sequence 4, Appl1
44	34.8	2.3	3468	3	US-08-459-444-4	Sequence 4, Appl1
45	34.8	2.3	3468	3	US-08-459-444-8	Sequence 8, Appl1

ALIGNMENTS

RESULT 1
US-08-773-368-2
; Sequence 2, Application US/08773368
; Patent No. 5856130
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Goll, Surya K.
; TITLE OF INVENTION: NOVEL HUMAN PATHOGENESIS-RELATED
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/773,368
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy RJ
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PR-0186 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 970 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: 1599164
; US-08-773-368-2
Query Match 5.0%; Score 74.6; DB 2; Length 970;
Best local Similarity 51.2%; Pred. No. 3.9e-11;

Matches 275; Conservative 0; Mismatches 226; Indels 36; Gaps 3;

```

QY 146 gagccatcccccagggaggaagagagatcctcatgtgcaacaagaagcttgggccc 205
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 88 GAGCCCTCACAGATGAGGAGAAACGTTTGATGTGAGCTGCACAACTTACCGGGCCC 147
QY 206 aggtgagagctcagagctcccaaatagagtaactgagagtgagagagagagag 265
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 148 AGGTATCCCGAGCGGCTCAGACATGCTGCACATGAGATGGAGAGAGAGTGGCCGCT 207
QY 266 ctgtctcagagctgggcaagctcagctcagctcagctcagctcagctcagctcag 325
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 208 TCGCCAGGCTTACGACGCGAGTTCGTTGCG-----GGCCACAAACAGAGCGCGGCG 261
QY 326 ccatgggagagagagctgggagctcagctcagctcagctcagctcagctcagctcag 385
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 262 GCGCGGCGGAGAAATCTGTTCGCTACACAGAGGAGGATGAGTGGCGCTGGCCATGG 321
QY 386 agtctgtatgagagagtgagagtgagagtgagagtgagagtgagagtgagagtgag 445
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 322 AGGATGGCACACGAGCTGAGACACTAACCTTACGCGCGGCGCT----- 369
QY 446 ggtgtcagagagagtgctcagctcagctcagctcagctcagctcagctcagctcag 505
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 370 -----GCAGCCGAGGCGCAGATGTGCGGCGCTACACGAGGTGGTATGGGCGCA 417
QY 506 ccacacaagaagagtggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 565
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 418 AGACAGAGAGAGATGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 471
QY 566 ttgggagagagagtggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 625
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 472 AGGAGACCAACATCGAATTAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 531
QY 626 aagccccctacaagaatgagctcagctcagctcagctcagctcagctcagctcag 682
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 532 AACGGCCCTACCAAGAGGAGGAGTCCGTGCTCCCAATGTCTCTGCTACACATGCA 588

```

RESULT 2

US-09-199-887-2

Sequence 2, Application US/09199887

Patent No. 6071874

GENERAL INFORMATION:

APPLICANT: Bandman, Olga

APPLICANT: Goli, Surya K.

TITLE OF INVENTION: NOVEL HUMAN PATHOGENESIS-RELATED

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESS: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09199,887

CLASSIFICATION:

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: 08/773,368

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy RJ

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0186 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 970 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

CLONE: 1599164

US-09-199-887-2

Query Match 5.0%; Score 74.6; DB 3; Length 970;
Best Local Similarity 51.2%; Pred. No. 3.9e-11;

Matches 275; Conservative 0; Mismatches 226; Indels 36; Gaps 3;

```

QY 146 gagccatcccccagggaggaagagagatcctcatgtgcaacaagaagcttgggccc 205
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 88 GAGCCCTCACAGATGAGGAGAAACGTTTGATGTGAGCTGCACAACTTACCGGGCCC 147
QY 206 aggtgagagctcagagctcccaaatagagtaactgagagtgagagagagagagag 265
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 148 AGGTATCCCGAGCGGCTCAGACATGCTGCACATGAGATGGAGAGAGAGTGGCCGCT 207
QY 266 ctgtctcagagctgggcaagctcagctcagctcagctcagctcagctcagctcag 325
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 208 TCGCCAGGCTTACGACGCGAGTTCGTTGCG-----GGCCACAAACAGAGCGCGGCG 261
QY 326 ccatgggagagagagctgggagctcagctcagctcagctcagctcagctcagctcag 385
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 262 GCGCGGCGGAGAAATCTGTTCGCTACACAGAGGAGGATGAGTGGCGCTGGCCATGG 321
QY 386 agtctgtatgagagagtgagagtgagagtgagagtgagagtgagagtgagagtgag 445
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 322 AGGATGGCACACGAGCTGAGACACTAACCTTACGCGCGGCGCT----- 369
QY 446 ggtgtcagagagagtgctcagctcagctcagctcagctcagctcagctcagctcag 505
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 370 -----GCAGCCGAGGCGCAGATGTGCGGCGCTACACGAGGTGGTATGGGCGCA 417
QY 506 ccacacaagaagagtggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 565
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 418 AGACAGAGAGAGATGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 471
QY 566 ttgggagagagagtggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 625
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 472 AGGAGACCAACATCGAATTAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 531
QY 626 aagccccctacaagaatgagctcagctcagctcagctcagctcagctcagctcag 682
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 532 AACGGCCCTACCAAGAGGAGGAGTCCGTGCTCCCAATGTCTCTGCTACACATGCA 588

```

RESULT 3

US-08-232-463-14/C

Sequence 14, Application US/08232463

Patent No. 5670367

GENERAL INFORMATION:

APPLICANT: DORNER, F.

APPLICANT: SCHRIFFLIN, F.

TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS

NUMBER OF SEQUENCES: 52

CORRESPONDENCE ADDRESS:

ADDRESS: Foley & Lardner

STREET: 1800 diagonal Road, Suite 500

CITY: Alexandria

STATE: VA

COUNTRY: USA

ZIP: 22313-0299

COMPUTER READABLE FORM:

GENERAL INFORMATION:
APPLICANT: Katz, L
APPLICANT: Donadio, S
APPLICANT: Mcalpine, J B
TITLE OF INVENTION: Recombinant DNA Method for Producing
TITLE OF INVENTION: Pyriminocin Analogs
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Edward H. Gorman
STREET: Abbott Laboratories D377/AP6D-2 One Abbott
STREET: Park Rd
CITY: Abbott Park
STATE: IL
COUNTRY: US
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/642,734C
FILING DATE: 17-JAN-91
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Danckers, Andreas M
REGISTRATION NUMBER: 32652
REFERENCE/DOCKET NUMBER: 4952.US.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-9196
TELEFAX: 708-938-2623
INFORMATION FOR SEO ID NO: 1 :
INFORMATION CHARACTERISTICS:

```

? topology: unknown
? molecule type: DNA (genomic)
? hypothetical: NO
? anti-sense: NO
? original source:
? organism: Saccharopolyspora erythraea
? strain: NRRL 2338
? feature:
?   name/key: misc_feature
?   location: 744..6659
?   other information: /function= "APPROXIMATE SPAN OF
?     OTHER INFORMATION: MODULE 1"
?     OTHER INFORMATION: /label= FUNCTION
? feature:
?   name/key: CDS
?   location: 744..11219
?   other information: /function= "gene= "eryA""
?     OTHER INFORMATION: /product= "OprL encoding modules 1 & 2 for
?       OTHER INFORMATION: 6-deoxyerythronolide B""
? feature:
?   name/key: misc_feature
?   location: 744..1868
?   other information: /function= "approximate span of
?     OTHER INFORMATION: acyltransferase domain 1 of module 1"
?     OTHER INFORMATION: acyltransferase domain 1 of module 1"
? feature:
?   name/key: misc_feature
?   location: 1998..2198
?   other information: /function= "approximate span of
?     OTHER INFORMATION: acyl carrier domain 1 of module 1"
?     OTHER INFORMATION: acyl carrier domain 1 of module 1"
? feature:
?   name/key: misc_feature
?   location: 2250..3626
?   other information: /function= "approximate span of
?     OTHER INFORMATION: beta-ketoacylACP synthase domain/module1"
?     OTHER INFORMATION: beta-ketoacylACP synthase domain/module1"
? feature:
?   name/key: misc_feature
?   location: 3831..4811

```

RESULT 5
 US-08-439-009A-1/c
 : Sequence 1, Application US/08439009A
 : Patent No. 6004787
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: Donadio, S
 :
 : APPLICANT: Katz, L
 :
 : APPLICANT: Malplene, J B
 :
 : TITLE OF INVENTION: Method of Directing Biosynthesis of
 :
 : TITLE OF INVENTION: Specific Polyketides
 :
 : NUMBER OF SEQUENCES: 27
 :
 : CORRESPONDENCE ADDRESS:
 :

```

1 ATTORNEY/AGENT INFORMATION:
2 NAME: Casuto, Dianne
3 REGISTRATION NUMBER: 40,943
4 REFERENCE/DOCKET NUMBER: 4952, US, D1
5 TELECOMMUNICATION INFORMATION:
6 TELEPHONE: 847-938-3137
7 INFORMATION FOR SEQ ID NO: 1:
8     SEQUENCE CHARACTERISTICS:
9         LENGTH: 11219 base pairs
10        TYPE: nucleic acid
11        STRANDEDNESS: double
12        TOPOLOGY: unknown
13        MOLECULE TYPE: DNA (genomic)
14        HYPOTHETICAL: NO
15        ANTI-SENSE: NO
16        ORIGINAL SOURCE:
17        ORGANISM: Saccharopolyspora erythraea
18        STRAIN: NRRL 2338
19
20 FEATURE:
21     NAME/KEY: misc_feature
22     LOCATION: 744..6559
23     OTHER INFORMATION: /function="APPROXIMATE SPAN OF
24     OTHER INFORMATION: MODULE 1"
25     OTHER INFORMATION: /label=FUNCTION
26
27 FEATURE:
28     NAME/KEY: CDS
29     LOCATION: 744..11219
30     OTHER INFORMATION: /function="gene" "eryA"
31     OTHER INFORMATION: /product="ORF1 encoding modules 1 & 2 for
32     OTHER INFORMATION: 6-deoxyerythronolide B"
33
34 FEATURE:
35     NAME/KEY: misc_feature
36     LOCATION: 744..1868
37     OTHER INFORMATION: /function="approximate span of
38     OTHER INFORMATION: acyltransferase domain 1 of module 1"
39
40 FEATURE:
41     NAME/KEY: misc_feature
42     LOCATION: 1998..2198
43     OTHER INFORMATION: /function="approximate span of
44     OTHER INFORMATION: acyl carrier domain 1 of module 1"
45
46 FEATURE:
47     NAME/KEY: misc_feature
48     LOCATION: 2250..3626
49     OTHER INFORMATION: /function="approximate span of
50     OTHER INFORMATION: beta-ketoacylACP synthase domain/module1"
51
52 FEATURE:
53     NAME/KEY: misc_feature
54     LOCATION: 3631..4811
55     OTHER INFORMATION: /function="approximate span of
56     OTHER INFORMATION: acyltransferase 2 domain of module 1"
57
58 FEATURE:
59     NAME/KEY: misc_feature
60     LOCATION: 5574..6125
61     OTHER INFORMATION: /function="approximate span of
62     OTHER INFORMATION: beta-ketoreductase domain of module 1"
63
64 NAME/KEY: misc_feature

```

Query Match	2.6%	Score 38.4	DB 4	Length 696
Best Local Similarity	48.2%	Pred. No. 0.28	Mismatches 116	Indels 0
Matches 108	Conservative	0	Gaps	0
22	gtaccccccttgggagctgtctgttcctgttcgtcggatcccaagagctaccctctggccaac	81		
DB				
77	gtcAACACCTCTCAGCTCGAGTTCCTCCGGAGTAAAGTCTCTACCTACGCCAGCGCCGC	136		
DB				
82	gtcaactctctttagagagagctgtctcagcaatatacagacaagagctctcaactccggatc	141		
QY				
DB	137	CACGTCGACGCGCACAGTCGACGAGCTGGCAGGGCGCGGCGTCAACATATGGTCCACACCTCCCGGAC	196	
QY				
142	cgcagagagcatcccccagagagagagagagagatctccctcagctgtgcacaaagcttcgg	201		
QY				
197	tggAAGCTCGGCTCCTGGGTACCAAGCAGGAGTACCTTGAGAGTCCCTCGACAGTATGCTG	256		
DB				
202	ggccaggtgtcagccttcagggccctcaacatgttgatgaatgacctg	245		
QY				
257	tgtgATATGTAAACGATAGCCGACGCCAGCCCTGCGGTTCTTTCCCG	300		
DB				

RESULT 7
 US-08-125-468-1/c
 ; Sequence 1, Application US/08125468
 ; Patent No. 5589385
 ;
 ; GENERAL INFORMATION:
 ;
 APPLICANT: Ryan, Michael J.
 ;
 APPLICANT: Lotvin, Nancy A.
 ;
 APPLICANT: Strathy, Jason A.
 ;
 APPLICANT: Fantini, Susan E.
 ;
 TITLE OF INVENTION: Cloning of the biosynthetic pathway for
 ; chlorotetracycline and tetracycline formation and cosmid
 ;
 TITLE OF INVENTION: useful therein
 ;
 TITLE OF INVENTION: useful therein
 ;
 NUMBER OF SEQUENCES: 1
 ;
 CORRESPONDENCE ADDRESS:
 ;
 ADDRESSEE: American Cyanamid Company
 ;
 STREET: One Cyanamid Plaza
 ;
 ;
 CITY: Wayne
 ;
 STATE: New Jersey
 ;
 COUNTRY: USA
 ;
 ZIP: 07470

RESULT 8
 US-08-474-933-1/C
 : Sequence 1, Application US/08474933
 : Patent No. 5866410
 :
 GENERAL INFORMATION:
 :
 APPLICANT: Ryan, Michael J.
 :
 APPLICANT: Lotvin, Jason A.
 :
 APPLICANT: Strathy, Nancy
 :
 TITLE OF INVENTION: Fantihi, Susan E.
 :
 TITLE OF INVENTION: Cloning of the biosynthetic pathway for
 :
 TITLE OF INVENTION: chlorotetracycline and tetracycline formation and cosmid
 :
 NUMBER OF SEQUENCES: 1
 :
 CORRESPONDENCE ADDRESS:
 :
 ADDRESSEE: American Cyanamid Company
 :
 STREET: One Cyanamid Plaza
 :
 CITY: Wayne
 :
 STATE: New Jersey
 :
 COUNTRY: USA
 :
 ZIP: 07470
 :
 COMPUTER READABLE FORM:
 :
 MEDIUM TYPE: Floppy disk
 :
 COMPUTER: IBM PC compatible
 :
 OPERATING SYSTEM: PC-DOS/MS-DOS
 :
 SOFTWARE: Patent In Release #1.0, Version #1.25
 :
 CURRENT APPLICATION DATA:
 :
 APPLICATION NUMBER: US/08/474,933
 :
 FILING DATE:
 :
 CLASSIFICATION: 435
 :
 PRIORITY APPLICATION DATA:
 :

	Query Match	Best Local Similarity	Score	DB 2:	Length	30001:
Matches	87, Conservative	51.88;	Pred. No. 1.2;	Mismatches	0;	Indels
						Gaps
Qy	160	gaggagaagaagagatccatcgtcgtgcacaaagcttcggaggccagtcgagccctcag	219			
Db	24305	gtggccggcgagggggaacttctacgcccggagaaactctggccggccgactgctccggac	219			
Qy	220	gctctcaacatgtgagtaacatgacctcgggatgcgaactcggagagctgtgtgtcaagctg	279			
Db	24245	accgaccacacatgacggcgctggctgcggccgacgagggggctccgccacacggggg	24246			
Qy	280	ggcagtcagtgagcattctcggagacgagggccaccagctgtgtgtgtgcc	327			
Db	24185	ggcgaccgcgacacacatgcggcgaacttctccggccggcgctcgtacggcc	24186			

RESULT 9
 US-08-159-784-1/c
 Sequence 1, Application US/08159784
 Patent No. 5643783
 GENERAL INFORMATION:
 APPLICANT: Bjorn R. Olsen
 TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF
 NUMBER OF SEQUENCES: 9
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Fish & Richardson
 STREET: 225 Franklin Street
 CITY: Boston
 STATE: Massachusetts
 COUNTRY: U.S.A.
 ZIP: 02110-2804
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
 COMPUTER: IBM PS/2 Model 50Z or 50SX
 OPERATING SYSTEM: MS-DOS (Version 5.0)
 SOFTWARE: WordPerfect (Version 5.1)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/159,784
 FILING DATE: December 1, 1993
 CLASSIFICATION: 530
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: John P. Freeman
 REGISTRATION NUMBER: 29,066
 REFERENCE/DOCKET NUMBER: 00246/170001
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 542-5070
 TELEFAX: (617) 542-8906
 TELEEX: 200154
 INFORMATION FOR SRO ID NO: 1:
 SEQUENCE CHARACTERISTICS:

```

; LENGTH: 4031
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-159-784-1

```

Query Match	2.5%;	Score 37.4;	DB 1;	Length 4031;	
Best Local Similarity	52.2%;	Pred. No. 1;			
Matches	83;	Mismatches	76;	Indels	0;
				Gaps	0

OY	402	ggtgaagacatacaactaccctcctaccggagcagtcaaccccttgcttcacagagatg	461
Db	1087	ggtcacccggttcacacattccttcttgaggagccatcttccctggtagcccttagagtc	1024
OY	462	cctcgagacctatgtgcagcaactacacagatatgttggcacaaccaagaatcgg	521
Db	1027	ccggggggtccttgtctcttcacaggacaggtttgttgaggctcttgaccatctggaccg	968
OY	522	tgtgctgtgaacacctgcgcggaagatgtaactgctggg	560
Db	967	caggaaccttgaggagccagctgggccaggtggagccttgag	929

RESULT 10
US-08-997-080-40
Application US/08997080

GENERAL INFORMATION:

APPLICANT: TAN, PAUL, L. J.
TITLE OF INVENTION: METHOD
NUMBER OF SEQUENCES: 194

COMPUTER READABLE FORM:

```

;
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/997,080
;

```

FILED ON: _____
ATTORNEY/AGENT INFORMATION: _____

REGISTRATION NUMBER: 37,007
REFERENCE/DOCKET NUMBER: 11000.1007
TELECOMMUNICATION INFORMATION:
305-369-0565

```

; TEL#X:
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1211 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
US-08-937-080-40

```

Query Match	2.5%;	Score 37;	DB 2;	Length 1211;
Best local Similarity	44.6%;	Pred. No. 0.84;		
Matches 145;	Conservative	0;	Mismatches 180;	Indels 0;
				Gaps 0;

[illegible]

RESULT 11
US-08-997-362-40
US/08997362

GENERAL INFORMATION:

APPLICANT: Hiyama, Jun
APPLICANT: Visser, Elizabeth
APPLICANT: Skinner, Margot
APPLICANT:

STREET: 2601 Elliott

COUNTRY: USA
ZIP: 98121
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette

```

; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Ver
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/997,362
;

```

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/873,970
 FILING DATE: June 12, 1997
 APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/705,347

ATTORNEY/AGENT INFORMATION:
NAME: Sleath, Janet
REGISTRATION NUMBER: 37,907
REFERENCE/DOCKET NUMBER: 11000.1002e2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-269-0565
TELEFAX: 206-269-0563
TELEX:
INFORMATION FOR SEQ ID NO: 40:
SEQUENCE INFORMATION:

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-08-997-362-40

Query Match
Best Local Similarity 2.5%; Score 37; DB 2; Length 1211;
Matches 145; Conservative 0; Mismatches 180; Indels 0; Gaps 0;

QY 119 acaagagtcaccccggttcgagagccatcccccagagagagagagatcc 178
DB 383 ACACCCAGCTGTACGCCGCCCGCTAACAGAGGCCCGCTGACCTACAACTGGGAGA 442
QY 179 tcatgtcacacaagagcttcgagccaggttgcaagcctcagagcctcaacttgataga 238
DB 443 CCTCTGTACCAAGAGCTCCCGGCTGCTGACAGCCACCGCGCTCAAGCCGACCG 502
QY 239 tgacctggatgacgaacttgagagctgtctgcaagcgttgagcagtcagtcctggg 298
DB 503 GCAGCGGCCCTGTGCTGTGATGGCGGCTGCGCCGCTGACCTGGCGACCTGGC 562
QY 299 agcagggcccccacagctgtctgtccatcgagcagaaacttgagcgctcactgggca 358
DB 563 ACCGGAGCACTTATCTACCGGGGCTCGATGTCGGCTTCTGAAACCCCTCCGAGGGCT 622
QY 359 ggtatgcttcgaggttcacatgtagctcgtgatatgagagtggaagactaacct 418
DB 623 GGTGGCGCTTCTGATCAACATCTCGATGGTGACCGCGGCTTCAAGCGGACGACA 682
QY 419 accctaccgagagtggaacc 443
DB 683 TGTGGGCGCAAGACGAGGGGATCCC 707

RESULT 12

US-08-873-970-40
Sequence 40, Application US/08873970
Patent No. 6001361

GENERAL INFORMATION:

APPLICANT: Tan, Paul
APPLICANT: Hiyma, Jun
APPLICANT: Visser, Elizabeth
APPLICANT: Skinner, Margot
APPLICANT: Prestidge, Ross
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Offices of Ann W. Speckman
STREET: 2601 Elliott Avenue, Suite 4185
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98121
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/873,970
FILING DATE:

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/705,347
FILING DATE: 29-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Sleeth, Janet
REGISTRATION NUMBER: 37,007
REFERENCE/DOCKET NUMBER: 11000.1002C1
TELECOMMUNICATION INFORMATION:

TELEPHONE: 206-269-0565
TELEFAX: 206-269-0563

INFORMATION FOR SEQ ID NO: 40:

SEQUENCE CHARACTERISTICS:
LENGTH: 1211 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-08-873-970-40

Query Match
Best Local Similarity 2.5%; Score 37; DB 3; Length 1211;
Matches 145; Conservative 0; Mismatches 180; Indels 0; Gaps 0;

QY 119 acaagagtcaccccggttcgagagccatcccccagagagagagatcc 178
DB 383 ACACCCAGCTGTACGCCGCCCGCTAACAGAGGCCCGCTGACCTACAACTGGGAGA 442
QY 179 tcatgtcacacaagagcttcgagccaggttgcaagcctcagagcctcaacttgataga 238
DB 443 CCTCTGTACCAAGAGCTCCCGGCTGCTGACAGCCACCGCGCTCAAGCCGACCG 502
QY 239 tgacctggatgacgaacttgagagctgtctgcaagcgttgagcagtcagtcctggg 298
DB 503 GCAGCGGCCCTGTGCTGTGATGGCGGCTGCGCCGCTGACCTGGCGACCTGGC 562
QY 299 agcagggcccccacagctgtctgtccatcgagcagaaacttgagcgctcactgggca 358
DB 563 ACCGGAGCACTTATCTACCGGGGCTCGATGTCGGCTTCTGAAACCCCTCCGAGGGCT 622
QY 359 ggtatgcttcgaggttcacatgtagctcgtgatatgagagtggaagactaacct 418
DB 623 GGTGGCGCTTCTGATCAACATCTCGATGGTGACCGCGGCTTCAAGCGGACGACA 682
QY 419 accctaccgagagtggaacc 443
DB 683 TGTGGGCGCAAGACGAGGGGATCCC 707

RESULT 13

US-09-095-855-40
Sequence 40, Application US/09095855
Patent No. 6160093

GENERAL INFORMATION:

APPLICANT: Tan, Paul
APPLICANT: Visser, Elizabeth
APPLICANT: Skinner, Margot
APPLICANT: Prestidge, Ross
TITLE OF INVENTION: Compounds and Methods for
NUMBER OF SEQUENCES: 208
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Offices of Ann W. Speckman
STREET: 2601 Elliott Avenue, Suite 4185
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98121
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/095,855
FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/705,347
FILING DATE: 29-AUG-1996

APPLICATION NUMBER: 08/873,970
FILING DATE: 12-JUN-1997
APPLICATION NUMBER: 08/997,362
FILING DATE: 23-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Sleath, Janet
REGISTRATION NUMBER: 37,007
REFERENCE/DOCKET NUMBER: 11000.1002c3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-269-0565
TELEFAX: 206-269-0563
TELEX:
INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 1211 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-09-095-855-40

Query Match 2.5%; Score 37; DB 4; Length 1211;
Best Local Similarity 44.6%; Pred. No. 0.84; Indels 0; Gaps 0;
Matches 145; Conservative 0; Mismatches 180;

OY 119 acaacgagctcactccgggtccgcagagccatcccccagagagagagagagatcc 178
DB 383 ACACCGACTGTACGCCGCCGCCCTTAACAAGGCGCCGACCGTACAACTGGGAGA 442
OY 179 tcatgtcgaacaagaagcttcgggccaagtgagcctcagcctcgaacatgagata 238
DB 443 CTTCTGACCCAGAGAGCTCCGCGCTGCGCAAGGCCAGCCGCGCTCAAGCCGACG 502
OY 239 tgacttgag 298
DB 503 GCAGGCGCCCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 562
OY 299 agcaggggcccacagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 358
DB 563 ACCCGGAGAGTTCATCTACGCGGCTCGATGTCCGCTTCTGAACCCCTCCGAGGCT 622
OY 359 ggtatgcctcgcggggttcctatgcaagtcctgtatgcaagtgaaagactac 418
DB 623 GGTGGCCGTTCTGATCAACATCTCGATGCGTGAACGCGCGCTTCAAGGCCGACG 682
OY 419 accctaccgag 443
DB 683 TGTGGGCAAGACGAGGAGATCCC 707

RESULT 14
US-08-705-347A-40
Sequence 40, Application US/08705347A
Patent No. 6284255
GENERAL INFORMATION:
APPLICANT: Tan, Paul
APPLICANT: Hiyma, Jun
APPLICANT: Visser, Elizabeth
APPLICANT: Skinner, Margot
APPLICANT: Scott, Linda
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND
TITLE OF INVENTION: DIAGNOSIS OF MYCOBACTERIAL INFECTIONS
NUMBER OF SEQUENCES: 55
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Speckman Picard PLLC
STREET: 2601 Elliott Avenue, Suite 4185
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98121
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/705,347A
FILING DATE: 28-AUG-1996
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Sleath, Janet
REGISTRATION NUMBER: 37,007
REFERENCE/DOCKET NUMBER: 11000.1002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206,269,0563
TELEFAX: 206,269,0563
INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 1211 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-705-347A-40

Query Match 2.5%; Score 37; DB 4; Length 1211;
Best Local Similarity 44.6%; Pred. No. 0.84; Indels 0; Gaps 0;
Matches 145; Conservative 0; Mismatches 180;

OY 119 acaacgagctcactccgggtccgcagagccatcccccagagagagagagagatcc 178
DB 383 ACACCGACTGTACGCCGCCGCCCTTAACAAGGCGCCGACCGTACAACTGGGAGA 442
OY 179 tcatgtcgaacaagaagcttcgggccaagtgagcctcagcctcgaacatgagata 238
DB 443 CTTCTGACCCAGAGAGCTCCGCGCTGCGCAAGGCCAGCCGCGCTCAAGCCGACG 502
OY 239 tgacttgag 298
DB 503 GCAGGCGCCCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 562
OY 299 agcaggggcccacagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 358
DB 563 ACCCGGAGAGTTCATCTACGCGGCTCGATGTCCGCTTCTGAACCCCTCCGAGGCT 622
OY 359 ggtatgcctcgcggggttcctatgcaagtcctgtatgcaagtgaaagactac 418
DB 623 GGTGGCCGTTCTGATCAACATCTCGATGCGTGAACGCGCGCTTCAAGGCCGACG 682
OY 419 accctaccgag 443
DB 683 TGTGGGCAAGACGAGGAGATCCC 707

RESULT 15
US-09-324-542-40
Sequence 40, Application US/09324542
Patent No. 6328978
GENERAL INFORMATION:
APPLICANT: Watson, James D.
APPLICANT: Tan, Paul L.J.
APPLICANT: Prestidge, Ross
TITLE OF INVENTION: Methods and Compounds for the Treatment
TITLE OF INVENTION: of Immunologically-Mediated Skin Disorders
FILE REFERENCE: 11000.1007c1
CURRENT APPLICATION NUMBER: US/09/324,542
CURRENT FILING DATE: 1999-06-02
EARLIER APPLICATION NUMBER: US 08/997,080
EARLIER FILING DATE: 1997-12-23
NUMBER OF SEQ ID NOS: 194
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 40
LENGTH: 1211
TYPE: DNA

! ORGANISM: Mycobacterium vaccae
US-09-324-542-40

Query Match 2.58; Score 37; DB 4; Length 1211;
Best Local Similarity 44.66; Pred. No. 0.84;
Matches 145; Conservative 0; Mismatches 180; Indels 0; Gaps 0;

```
QY 119 acacgagctcactcccggtccgcagagcattccccaaggaggaagatcc 178
    ||| ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 383 acacgagctgtacgcccccgccgtacaagaaggcccgactgaacttaagtgga 442
QY 179 tcatgtgcacaaagcttcggggccagggtcagcctcaggcctcaaatgagtlaca 238
    | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 443 ccttctgaccacagagctcccggtctgtgcaggccaaccgcggtcaagccgaccg 502
QY 239 tgacctggagtgacgaactggagaagctctgtcagcagctggccagtcagtcattgg 298
    | | | | | | | | | | | | | | | | | | | | | |
Db 503 gcagcgccctgtcggtctgtcgatggcggttcggccgctgaacctggcagacctggc 562
QY 299 agcagcgcccaaccagctcgtcgtgtccatcgggcaagaacctggggcgctcactgggca 358
    | | | | | | | | | | | | | | | | | | | | | |
Db 563 acccgagagcagtcactacgcgggtctcgatgtccggtctcctgaacccctcggagggt 622
QY 359 ggtatcgctctccgggttcacatgtcagtcctcgtgtatgaagagtgaaagactaacct 418
    ||| | | | | | | | | | | | | | | | | | | |
Db 623 ggtggcggttcctgataacatctgatgggtgaacggcggtcttaaggcgagcagaca 682
QY 419 accctaccggagcgagtgcaacc 443
    | | | | | | | | | | | | | | | | | |
Db 683 tgtgggcaagaaccgaggatccc 707
```

Search completed: May 25, 2002, 13:23:33
Job time: 10224 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 25, 2002, 13:23:34 ; Search time 61 Seconds

(without alignments)
199,009 Million cell updates/sec

Title: US-09-667-380a-2

Perfect score: 2759 1 MSCVLGCVLPGLFLVCGS.....SESLGTPROGKAPRIRAVRQ 497

Sequence: 1 MSCVLGCVLPGLFLVCGS.....SESLGTPROGKAPRIRAVRQ 497

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues 231628

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	355	12.9	266	2	US-08-773-368-4
2	355	12.9	266	3	US-09-199-887-4
3	352	12.8	219	2	US-08-773-368-3
4	352	12.8	219	3	US-09-199-887-3
5	322	11.7	270	2	US-08-773-368-1
6	322	11.7	270	3	US-09-199-887-1
7	273	9.9	220	5	PCT-US96-07709-30
8	273	9.9	248	5	PCT-US96-07709-25
9	269	9.7	204	5	PCT-US96-07709-33
10	256	9.3	204	1	US-08-419-414-9
11	256	9.3	204	1	US-08-614-935-2
12	256	9.3	204	3	US-09-130-287-2
13	244.5	8.9	440	1	US-07-930-686-12
14	244.5	8.9	440	2	US-08-460-998-12
15	240	8.7	204	1	US-08-614-935-1
16	240	8.7	204	3	US-09-130-287-1
17	237	8.6	203	1	US-08-419-414-8
18	237	8.6	203	3	US-08-614-935-3
19	237	8.6	203	1	US-09-130-287-3
20	236.5	8.6	424	1	US-08-419-414-2
21	227	8.2	151	1	US-08-614-935-28
22	227	8.2	151	3	US-09-130-287-28
23	226	8.2	205	5	US-08-450-944-5
24	226	8.2	205	5	PCT-US96-07709-5
25	226	8.2	221	5	US-08-450-944-2
26	226	8.2	221	5	PCT-US96-07709-2
27	225.5	8.2	137	2	US-07-857-224B-110

28	222.5	8.1	205	1	US-08-419-414-10	Sequence 10, Appl
29	222.5	8.1	205	1	US-08-614-935-6	Sequence 6, Appl
30	222.5	8.1	205	1	US-08-614-935-7	Sequence 7, Appl
31	222.5	8.1	205	3	US-09-130-287-6	Sequence 6, Appl
32	222.5	8.1	205	3	US-09-130-287-7	Sequence 7, Appl
33	222	8.0	454	1	US-07-930-686-10	Sequence 10, Appl
34	222	8.0	454	2	US-08-460-998-10	Sequence 10, Appl
35	217.5	7.9	205	3	US-09-130-287-4	Sequence 4, Appl
36	217.5	7.9	204	3	US-08-614-935-5	Sequence 5, Appl
37	217	7.9	204	3	US-09-130-287-5	Sequence 5, Appl
38	217	7.9	135	2	US-07-857-224B-103	Sequence 103, App
39	216	7.8	166	1	US-08-614-935-30	Sequence 30, Appl
40	216	7.8	166	3	US-09-130-287-30	Sequence 29, Appl
41	216	7.8	150	3	US-08-614-935-29	Sequence 29, Appl
42	212.5	7.7	150	3	US-09-130-287-29	Sequence 112, App
43	212.5	7.7	138	2	US-07-857-224B-111	Sequence 111, App
44	211	7.6	137	2	US-07-857-224B-111	Sequence 111, App
45	210.5	7.6	137	2	US-07-857-224B-111	Sequence 111, App

ALIGNMENTS

RESULT 1
US-08-773-368-4
; Sequence 4, Application US/08773368
; Patent No. 5856130
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Golli, Surya K.
; TITLE OF INVENTION: NOVEL HUMAN PATHOGENESIS-RELATED
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/773,368
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy RJ
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0186 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 266 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1030053
; US-08-773-368-4
Query Match 12.9%; Score 355; DB 2; Length 266;

QY 94 ASOCIMHG-----PTSL--LVSIGONIGAHMGCRSPGFHVOS-----WYDEVKDYTP 141
DB 62 ASNCFSHNTRLRKPKHLHPNTSLGENI---W-TGSPVPIFSVSSALITWYDEIDQYNF- 116
QY 142 YPSECNWPCRGSPMCTHYTOIWMATNKIGCAVNTCKRMVGEWENAVYFCVNS 201
DB 117 -----KTRICKKVCCHTYQVWADSYKVCACVFCPCVSGF-DALSNAGHFICNTG 166
QY 202 PKGN---WIGEAPYKNGRCPSECPSPSYGSGCRNNLCYRE 237
DB 167 PGGNPTW---PYKRGATCSACPNN--DKCLDNLGVND 199

RESULT 4

US-09-199-887-3
Sequence 3, Application US/09199887
Patent No. 6071874
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Goll, Surya K.
TITLE OF INVENTION: NOVEL HUMAN PATHOGENESIS-RELATED
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/199,887
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/773,368
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy RJ
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0186 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 219 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 847722
US-09-199-887-3

Query Match 12.8%; Score 352; DB 3; Length 219;
Best Local Similarity 37.0%; Pred. No. 1.3e-24;
Matches 81; Conservative 31; Mismatches 67; Indels 40; Gaps 12;

QY 35 LSKYOHNESSRVRRAIPRD-KEIILMLHNKLRGOVQPOASNMEYMTWDELEKSAANV 93
DB 5 VSNSSH---ANTLPDENEDFKDCVRINKRSEYKPTASDMLITWDPALAAQAKAN 61
QY 94 ASOCIMHG-----PTSL--LVSIGONIGAHMGCRSPGFHVOS-----WYDEVKDYTP 141

DB 62 ASNCFSHNTRLRKPKHLHPNTSLGENI---W-TGSPVPIFSVSSALITWYDEIDQYNF- 116
QY 142 YPSECNWPCRGSPMCTHYTOIWMATNKIGCAVNTCKRMVGEWENAVYFCVNS 201
DB 117 -----KTRICKKVCCHTYQVWADSYKVCACVFCPCVSGF-DALSNAGHFICNTG 166
QY 202 PKGN---WIGEAPYKNGRCPSECPSPSYGSGCRNNLCYRE 237
DB 167 PGGNPTW---PYKRGATCSACPNN--DKCLDNLGVND 199

RESULT 5

US-08-773-368-1
Sequence 1, Application US/08773368
Patent No. 5856130
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Goll, Surya K.
TITLE OF INVENTION: NOVEL HUMAN PATHOGENESIS-RELATED
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/773,368
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy RJ
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0186 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 270 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: 1599164
US-08-773-368-1

Query Match 11.7%; Score 322; DB 2; Length 270;
Best Local Similarity 35.6%; Pred. No. 1e-21;
Matches 84; Conservative 27; Mismatches 105; Indels 20; Gaps 8;

QY 50 AIPREDKEIILMLHNKLRGOVQPOASNMEYMTWDELEKSAANVASOCIMHGPTSLVS 109
DB 27 ALTDEKRLMVLNHLNLYRAQVSPFASDMLHMRWDEBLAFAKAYARQ--XRXGHNKERGR 84
QY 110 IQONIGAHMGCRSPGFHVOSYDEVKDYTPYPSNCPWPCRGSPMCTHYTOIWMAT 169
DB 85 RGENLFAITDEGMDVPLAEWHEHERHYNLS-AAITCSP-----GOMCGHYTYVWAK 136
QY 170 TNKIGCAVNTCKRMVGEWENAVYFCVNSPKGNWIGEADYKNGRCPSECPSPSYGSGC 229

Db 137 TERIGCGSHFEKIQ--GVEETNIELLYCANEPPGNVKGKRPQEGTSCCSPSGY--HC 192
QY 230 RNNLCYRETTYPKPEDEMEVEFAPPEENHVLQPRVW--RPRPKTSAYN 282
Db 193 KNSLC--EPISPDADQLPYLVEAFSPRATED--SDSRKMGAGPDKPSVSGLN 245

RESULT 6

US-09-199-887-1
Sequence 1, Application US/09199887
Patent No. 6071874
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Goll, Surya K.
TITLE OF INVENTION: NOVEL HUMAN PATHOGENESIS-RELATED
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/199,887
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/773,368
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy RJ
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0186 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 270 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: 1595164
US-09-199-887-1

Query Match 11.7%; Score 322; DB 3; Length 270;
Best Local Similarity 35.6%; Pred. No. 1e-21;
Matches 84; Conservative 27; Mismatches 105; Indels 20; Gaps 8;

QY 50 AIRDEKEEILMLHNLKLGQVPOASNMETMDDELEKSAAMASQCWEHGPTSLVLS 109
Db 27 ALDEDEKRLVDELHNLRYQVSTASDMLHMRDELAFAAAYARQ--XRGHNERGR 84
QY 110 ICONLGAHNGRYSRPGFHVQSWYDEVKDYTPYSPHCNFWCEPSCGPMCTHYTOTVWAT 169
Db 85 RGENLEFAITDEGDVPLAMEWHHEREHYNLS-AATCSP-----GQMGHITVYVMAK 136
QY 170 TNGICAVNTCKRMVWGVWENAVFYVCNYSPKGNWIGEAPYKNGRSPCSPSYGSGC 229
Db 137 TRIGCGSHFEKIQ--GVEETNIELLYCANEPPGNVKGKRPQEGTSCCSPSGY--HC 192
QY 230 RNNLCYRETTYPKPEDEMEVEFAPPEENHVLQPRVW--RPRPKTSAYN 282

Db 193 KNSLC--EPISPDADQLPYLVEAFSPRATED--SDSRKMGAGPDKPSVSGLN 245

RESULT 7

PCT-US96-07709-30
Sequence 30, Application PC/TUS9607709
GENERAL INFORMATION:
APPLICANT: Tripp, Cynthia A.
APPLICANT: Wisniewski, Nancy
TITLE OF INVENTION: PARASITIC HELMINTH VENOM ALLERGEN
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Rosa & McIntosh
STREET: 1700 Lincoln St., Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/07709
FILING DATE: 23-MAY-1996
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-30-PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 220 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US96-07709-30

Query Match 9.9%; Score 273; DB 5; Length 220;
Best Local Similarity 29.5%; Pred. No. 2.5e-17;
Matches 70; Conservative 36; Mismatches 85; Indels 46; Gaps 10;

QY 16 LVCGSGYLLP--NWLLEELLSKYQHNSHSRVRRAIPREDKEILMLHNLKLR---GQV 70
Db 10 IYVAAYGVNCPGCKLTALER-----KRIYQNNKTRSDILNKLNRNGTY 55
QY 71 OPOASNMETMDDELEKSAAMASQCWEHGPTSLVLSIGONLGAHW-----GRYRSP 124
Db 56 MPRGNMELRLWDCKLESSAOMANOCIFGHSPPROQREGVEENYAYVSSVSGELKTA 115
QY 125 GFHV-QSWYDEVKDYTPYSPHCNFWCEPSCGPMCTHYTOTVWATNKKICAVNT-C-- 180
Db 116 GTDACKSMWSSELPKLYENPNPNKMTW---KYAGQVLFHTQMAMGKTYKIGCGVATOCODG 172
QY 161 -RKMTVMEVEWENAVFYVCNYSPKGNWIGEAPYKNGRSPCSPSYGSGC--RNNLC 234
Db 173 GRTLV-----ICHYSPGNNMVGELYTGKNDCKYDKCYTKKCLSLSGCLC 218

RESULT 8

PCT-US96-07709-25
Sequence 25, Application PC/TUS9607709
GENERAL INFORMATION:
APPLICANT: Tripp, Cynthia A.
APPLICANT: Wisniewski, Nancy
TITLE OF INVENTION: PARASITIC HELMINTH VENOM ALLERGEN
TITLE OF INVENTION: ANTIGEN 5-LIKE GENES AND PROTEINS

```

1      OPERATING SYSTEM:  PC-DOS/MS-DOS      PC-DOS/MS-DOS
2      SOFTWARE:  PatentIn Release #1.0, Version #1.25
3      CURRENT APPLICATION DATA:
4      APPLICATION NUMBER:  PCT/US96/07709
5      FILING DATE:  23-MAY-1996
6      CLASSIFICATION:
7      ATTORNEY/AGENT INFORMATION:
8      NAME:  Connell, Gary J.
9      REGISTRATION NUMBER:  32,020
10     REFERENCE/DOCKET NUMBER:  2618-30-PCT
11     TELECOMMUNICATION INFORMATION:
12     TELEPHONE:  (303) 863-9700
13     TELEFAX:  (303) 863-0223
14     INFORMATION FOR SEQ ID NO:  33:
15     SEQUENCE CHARACTERISTICS:
16     ?     LENGTH:  204 amino acids
17     ?     TYPE:  amino acid
18     ?     TOPOLOGY:  linear
19     ?     MOLECULE TYPE:  protein
20     ?     PCT-US96-07709-33

```

REFERENCE/DOCKET NUMBER: YU113
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404) 873-8795
TELEFAX: (404) 873-8795
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 204 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Vesputia vulgaris
US-08-419-414-9

Query Match 9.3%; Score 256; DB 1; Length 204;
Best Local Similarity 30.5%; Pred. No. 8.1e-16;
Matches 71; Conservative 33; Mismatches 73; Indels 56; Gaps 12;

QY 1 MSCVIGVIRPLGLFLVCGSGYLTPNVLLELLSKYQHNSHSHVRAIRPREDEEIL 60
DB 6 IKCLKGV-----HTAC-KYSLKPNCG--NKVYVY-----GLTKQEKDIL 45
QY 61 MLNKLRGV-----OPASNMEYTWDELEKSAAMASOCIEHGFTSLIV 108
DB 46 KEHNDROKRIAGLETGRNPGPQPPAKNNKLVNDELAYVAQVWANCQYGHDTCDVA 105
QY 109 --SIGONTL--GAHNGRRSPGFHVSQYDEKDYTYYPSECPNCPERCSP--MCT 160
DB 106 KYQVGNVALGSIATAKIDDPKLVKMEDEVKDY-----NP--KKFSGNDLKTG 155
QY 161 HTQIVMATNTKIGCAVNTCRKMTWGEVMEVNAVYFCNTSPKGNMGEAPYK 213
DB 156 HTQIVMATNTKEVCG-----SIKTIQEKMKH-KLVLCNTPGSGNEMNEELYQ 202

RESULT 11
US-08-614-935-2

Sequence 2, Application US/08614935
Patent No. 5804201

GENERAL INFORMATION:

APPLICANT: King, Te P.
TITLE OF INVENTION: IMMUNOMODULATORY PEPTIDES OF VESPID
NUMBER OF SEQUENCES: 81
CORRESPONDENCE ADDRESS:
ADDRESSEE: David A. Jackson, Esq.
STREET: 411 Hackensack Ave, Continental Plaza, 4th
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/614,935
FILING DATE: 11-MAR-1996
CLASSIFICATION: 436

ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-156
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 204 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE:
ORIGINAL SOURCE:
ORGANISM: Vesputia vulgaris
US-08-614-935-2

Query Match 9.3%; Score 256; DB 1; Length 204;
Best Local Similarity 30.5%; Pred. No. 8.1e-16;
Matches 71; Conservative 33; Mismatches 73; Indels 56; Gaps 12;

QY 1 MSCVIGVIRPLGLFLVCGSGYLTPNVLLELLSKYQHNSHSHVRAIRPREDEEIL 60
DB 6 IKCLKGV-----HTAC-KYSLKPNCG--NKVYVY-----GLTKQEKDIL 45
QY 61 MLNKLRGV-----OPASNMEYTWDELEKSAAMASOCIEHGFTSLIV 108
DB 46 KEHNDROKRIAGLETGRNPGPQPPAKNNKLVNDELAYVAQVWANCQYGHDTCDVA 105
QY 109 --SIGONTL--GAHNGRRSPGFHVSQYDEKDYTYYPSECPNCPERCSP--MCT 160
DB 106 KYQVGNVALGSIATAKIDDPKLVKMEDEVKDY-----NP--KKFSGNDLKTG 155
QY 161 HTQIVMATNTKIGCAVNTCRKMTWGEVMEVNAVYFCNTSPKGNMGEAPYK 213
DB 156 HTQIVMATNTKEVCG-----SIKTIQEKMKH-KLVLCNTPGSGNEMNEELYQ 202

RESULT 12
US-09-130-287-2

Sequence 2, Application US/09130287
Patent No. 6106844

GENERAL INFORMATION:

APPLICANT: King, Te P.
TITLE OF INVENTION: IMMUNOMODULATORY PEPTIDES OF VESPID
NUMBER OF SEQUENCES: 81
CORRESPONDENCE ADDRESS:
ADDRESSEE: David A. Jackson, Esq.
STREET: 411 Hackensack Ave, Continental Plaza, 4th
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/130,287
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/614,935
FILING DATE: 11-MAR-1996
CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-156 DIV
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
 LENGTH: 204 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHEetical: NO
 ANTI-SENSE: NO
 FRAGMENT TYPE:
 ORGANISM: Vespula vulgaris
 US-09-130-287-2

Query Match 9.3%; Score 256; DB 3; Length 204;
 Best Local Similarity 30.5%; Pred. No. 8.1e-16;
 Matches 71; Conservative 33; Mismatches 73; Indels 56; Gaps 12;

QY 1 MSCVLGVIPILGLFLVCGSGYLLPNTLLLELLSKYOHNSHVRRAIPREDKEILL 60
 DB 6 IKCLKGV-----HTAC-KYGLKPNCG--NKVVS-----GLTKOEKODILL 45
 QY 61 MHNLRGV-----QPQASNMEYMTWDELKSAANAASOCIWEHPTSLV 108
 DB 46 KEHNEFRQIARLETRGNPGPPAKNKNLVWDELAYVAQVMANQCYGHDTCRDVA 105
 QY 109 --SIGONT--GAHWGRYRSPGFHVOSWYDEVKDYTPPSECPNCPERCSCP--MCT 160
 DB 106 KVOGQONVALISTAKAYDPYKLVLMWDEYKDY-----NF--KKFSNDLKLNG 155
 QY 161 HTQIVWATNTKICAVNTCRKMTWGEVWENAVFYCNYSKGNMIGEARPYK 213
 DB 156 HXTQWVWANTKEVCGC-----SIKYIOEKMKH-KYLVCMYSGNFMNELYQ 202

RESULT 13

US-07-930-686-12
 ; Sequence 12, Application US/07930686
 ; Patent No. 5525508

GENERAL INFORMATION:
 APPLICANT: Sharp, Phillip J
 APPLICANT: Magland, Barry M
 APPLICANT: Cobon, Gary S
 TITLE OF INVENTION: Nematode Vaccine
 NUMBER OF SEQUENCES: 12
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Foley and Lardner
 STREET: suite 500, 1800 Diagonal Road
 CITY: Alexandria
 STATE: Virginia
 COUNTRY: United States of America
 ZIP: 22313-0299
 COMPUTER READABLE FORM:
 MEDIUM TYPE: floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/930,686
 FILING DATE: 19921006
 CLASSIFICATION: 514
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: AU PK4486
 FILING DATE: 06-FEB-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/AU92/00040
 FILING DATE: 06-FEB-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Bent, Stephen A
 REGISTRATION NUMBER: 29,768
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 836-9300
 TELEFAX: (703) 683-4109

TELEX: 899149
 ; INFORMATION FOR SEQ ID NO: 12:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 440 amino acids
 ; TYPE: AMINO ACID
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-07-930-686-12

Query Match 8.9%; Score 244.5; DB 1; Length 440;
 Best Local Similarity 29.7%; Pred. No. 2.8e-14;
 Matches 73; Conservative 32; Mismatches 86; Indels 55; Gaps 10;

QY 39 OHNSHVRRAIPREDKEILLMHNLRGVQPOASNMEYMTWDELKSAANAASOCI 98
 DB 32 KHEHYRSIIAKG-----QAKNKL-GGFAPKARMLKYDYDEVENATVAAYKECK 80
 QY 99 WEHPTSLVSTIGONT--GAHWGRYRSPGFHVOSWYDEVKDYTPPSECPNCPERC 154
 DB 81 FEHDPEORRYWQONLMMLGTYNYSKTESAKLSVOAMWELKMGVDPDENILTMVEYFDG 140
 QY 155 SGPMCTHYQIVWATNTKICAVNTCRKMTWGEVWENAVFYCNYSKGNMIGEARPYKN 214
 DB 141 VG-----HTQVWVWANTKEVCGC-----VACEYINPAGNRINHYIYDI 186
 QY 215 GRPCS--ECPPSYGSGCRNNLCYRETY-----TPKPEDEMEVETAP----- 256
 DB 187 GDCPTDEDC-----QCCTCSCSDKALCIPGYTVMPTTE--KPTTPKLYHPGGM 238
 QY 257 IPEENH 262
 DB 239 CPENNN 244

RESULT 14

US-08-460-998-12
 ; Sequence 12, Application US/08460998
 ; Patent No. 5942413

GENERAL INFORMATION:
 APPLICANT: Sharp, Phillip J
 APPLICANT: Magland, Barry M
 APPLICANT: Cobon, Gary S
 TITLE OF INVENTION: Nematode Vaccine
 NUMBER OF SEQUENCES: 12
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Foley and Lardner
 STREET: suite 500, 3000 K Street, NW
 CITY: Washington
 STATE: DC
 COUNTRY: United States of America
 ZIP: 20007-5109
 COMPUTER READABLE FORM:
 MEDIUM TYPE: floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/460,998
 FILING DATE: 05-JUN-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/930,686
 FILING DATE: 06-OCT-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: AU PK4486
 FILING DATE: 06-FEB-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/AU92/00040
 FILING DATE: 06-FEB-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Bent, Stephen A
 REGISTRATION NUMBER: 29,768
 TELECOMMUNICATION INFORMATION:

Search completed: May 25, 2002, 14:28:50
Job time: 3916 sec

C:Accession: J04131
 R:Murphy, E.V.; Zhang, Y.; Zhu, W.; Biggs, J.
 Gene 159, 131-135, 1995
 A:Title: The human glioma pathogenesis-related protein is structurally related to plant
 A:Reference number: J04131; MUID:9531646
 A:Molecule type: mRNA
 A:Residues: 1-219 <MUI>
 A:Cross-references: GB:016307; NID:91100927; PIDN:AAA2731.1; PID:g847722
 A:Experimental source: brain tumor
 C:Genetics:
 A:Gene: GDB:GILPR
 A:Cross-references: GDB:683195
 C:Superfamily: yellowjacket venom allergen antigen 5
 C:Keywords: brain

Query Match

Best Local Similarity 12.8%; Score 352; DB 2; Length 219;
 Matches 81; Conservative 31; Mismatches 67; Indels 40; Gaps 12;

QY 35 LSKYOHESHRYRAIPRED-KEEILMLHNKLGVOVQASNMETWDELKSAAM 93
 DB 5 VSNYSHT---ANILPDIEDNEPINDCVRIHNRKSEVKTASDMLYMTDPAIAQAKAM 61
 QY 94 ASQCIWEG-----PTSL---LVISQNLGAMGRYRSPGFHYOS---WYDEVKDYTP 141
 DB 62 ASNCOFSHNTRLKPPHKLHFNFTSLGENT---W-TGSVPIFSVSAITWYDEIDYNF- 116
 QY 142 YPSECNWCPERSGPMCTHYTOIYMATNKGCAVNTCKMTWCEWENAVYFCVNCYS 201
 DB 117 -----KTRICKVCGHYTOYVMAVDSYKCAVQFCPKVSGP-DALSGAIFICNG 166
 QY 202 PKNN---WIGFAPIYKNGPCSECPSPYSGSCRNNUCYRE 237
 DB 167 PEGNPTW-----PYKRGATSCACPNV-DKCLDNLCVND 199

RESULT 3
 A24609
 A:Accession: J04131
 N:Alternative names: sperm-coating glycoprotein precursor - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 30-Jun-1988 #sequence-revision 30-Jun-1988 #text-change 29-Sep-1999
 R:Chastet, N.J.; Joseph, D.R.; Wilson, E.M.; French, F.S.
 Molecule type: preliminary
 A:Title: Molecular cloning of complementary deoxyribonucleic acid for an androgen-regulated
 A:Reference number: A40918; MUID:89039913
 A:Accession: A40918
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-246 <CHN>
 A:Cross-references: GB:M31173; NID:9202772; PIDN:AA859716.1; PID:9202773
 R:Brooks, D.E.; Means, A.R.; Wright, E.J.; Singh, S.P.; Tlaver, K.K.
 A:Title: Molecular cloning of the cDNA for androgen-dependent sperm-coating glycoprotein
 A:Reference number: A24609; MUID:87053995
 A:Accession: A24609
 A:Residues: 1-246 <BRN>
 C:Superfamily: cysteine-rich secretory protein 1
 C:Keywords: cysteine-rich secretory protein 1
 F:1-19/Domain: signal sequence
 F:20-246/Product: acidic epididymal glycoprotein

Query Match
 Best Local Similarity 10.9%; Score 301.5; DB 2; Length 246;
 Matches 81; Conservative 51; Mismatches 90; Indels 81; Gaps 15;
 QY 11 LGLELVGSGGILPNVTLLELLSKYOHN-ESHSHVRAIRPREDKEEILMLHNKLGQ 69

DB 5 LVLEFLAA-----VLP-PSLLDPTDDEWRDLENSTKLVS-----QEEIKHNDLQRT 54
 QY 70 VPOASNMETWDELKSAAMASOCIEWEGP-----TSLLVSGNT-----GAHGRY 121
 DB 55 VSPSGDILLVEMDHDVAVNAQKAMNCTYHSPLOHRTTTLKGENLFMANYPASMSV 114
 QY 122 RSPGFHYOSMYDEVKDYTPYPSECNWCPERSGPMCTHYTOIYMATNKGCAVNTCR 181
 DB 115 -----IDWIDESLIDFVFGG-----PKRV-GVAKGHITOVVNSFLVACVACCP 160
 QY 182 KMTWGEWENAVYFCVNCSPKGNMIGE--APYKNGRPGCEPSPYSGSCRNNUCYRE 239
 DB 161 DQPL-----KFFVCHYCPGAGNVGRVSPYTEGEFCDSCP-----GNCEDGLCTNSCE 209
 QY 240 YTPKPEIDENKNEVEYAPIDENHVMLOPRVMPRTKPKTSAVNVMTQVYRCDTR-MKDRG 298
 DB 210 Y-----EDNT-----SNCGDLKMYSCDDPLKEGC 235
 QY 299 KGS 301
 DB 236 RAS 238

RESULT 4

A49202
 A:Accession: J04131
 N:Alternative names: CRISP-1
 C:Species: Mus musculus (house mouse)
 C:Date: 19-Dec-1993 #sequence-revision 18-Nov-1994 #text-change 29-Sep-1999
 R:Haendler, B.; Kratzschmar, J.; Theuring, F.; Schlemming, W.D.
 A:Title: Transcripts for CRISP-1, a cysteine-rich secretory protein-1 (CRISP-1; DE/EC) and the
 A:Reference number: A49202; MUID:93307144
 A:Accession: A49202
 A:Status: preliminary
 A:Molecule type: nucleic acid
 A:Cross-references: GB:I05559; NID:9309190; PIDN:AAA37460.1; PID:g309191
 A:Experimental source: NMRI, epididymis, salivary gland
 A:Note: sequence extracted from NCBI backbone (NCBI:134675, NCBI:134676)
 C:Superfamily: cysteine-rich secretory protein 1

Query Match

Best Local Similarity 10.8%; Score 298; DB 2; Length 244;
 Matches 77; Conservative 35; Mismatches 87; Indels 42; Gaps 11;

QY 11 LGLELVGSGGILPNVTLLELLSKYOHNESHSHVRAIRPREDKEEILMLHNKLGQ 70
 DB 5 LVLEFLAA-----VLP-PSLLD-SSQENRLEKLTITKSV-----QEEIKHNDLQRT 54
 QY 71 OPOASNMETWDELKSAAMASOCIEWEGP-----TSLLVSGNT-----GAHGRY 122
 DB 54 SPSSGDLKMEWMDVAVNAQKAMNCTYHSPLOHRTTTLKGENLFMANYPASMSV 114
 QY 123 SPGHVOSMYDEVKDYTPYPSECNWCPERSGPMCTHYTOIYMATNKGCAVNTCR 181
 DB 112 -----AIOGVNEXKDLIT-----DVGPKQDPSYVG-----HYTOVYNSIFVOACVACCP 159
 QY 183 KMTWGEWENAVYFCVNCSPKGNMIGE--APYKNGRPGCEPSPYSGSCRNNUCYRE 239
 DB 160 NPL-----RYTVCHYCPVNGYGRVLTPTTAGECACSCPHCEDGLCTNSCHDEK 212
 QY 241 T 241
 DB 213 T 213

RESULT 5

S68691
 A:Accession: J04131
 N:Alternative names: CRISP-1
 C:Species: Mus musculus (house mouse)
 C:Date: 19-Dec-1993 #sequence-revision 18-Nov-1994 #text-change 29-Sep-1999
 R:Haendler, B.; Kratzschmar, J.; Theuring, F.; Schlemming, W.D.
 A:Title: Transcripts for CRISP-1, a cysteine-rich secretory protein-1 (CRISP-1; DE/EC) and the
 A:Reference number: A49202; MUID:93307144
 A:Accession: A49202
 A:Status: preliminary
 A:Molecule type: nucleic acid
 A:Cross-references: GB:I05559; NID:9309190; PIDN:AAA37460.1; PID:g309191
 A:Experimental source: NMRI, epididymis, salivary gland
 A:Note: sequence extracted from NCBI backbone (NCBI:134675, NCBI:134676)
 C:Superfamily: cysteine-rich secretory protein 1

neutrophil granules matrix glycoprotein SGP28 precursor - human

C:Species: Homo sapiens (man)
 C:Date: 15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 17-Nov-2000
 C:Accession: S68691; S74313; S68683
 R:Kjeldsen, L.; Cowland, J.B.; Johnsen, A.H.; Borregaard, N.
 FEBS Lett. 380, 246-250, 1996
 A:Title: SGP28, a novel matrix glycoprotein in specific granules of human neutrophils
 A:Reference number: S68691; MUID:96186934
 A:Accession: S68691
 A:Molecule type: mRNA
 A:Residues: 1-245 <KJE>
 A:Cross-references: EMBL:X94323; NID:g1213612; PIDN:CAA63984.1; PID:g1213613
 A:Accession: S74313
 A:Molecule type: protein
 A:Residues: 33-83;96-143;165-217;221-226 <KJL>
 R:Kraetzschmar, J.; Haendler, B.; Eberspaecher, U.; Roosterman, D.; Donner, P.; Schleunig, J.; Blochem, 236, 827-836, 1996
 Eur. J. Biochem. 236, 827-836, 1996
 A:Title: The human cysteine-rich secretory protein (CRISP) family. Primary structure and
 A:Reference number: S68681; MUID:96270732
 A:Accession: S68683
 A:Molecule type: mRNA
 A:Status: preliminary
 A:Residues: 1-105; 'S', 107-245 <KRA>
 A:Cross-references: EMBL:X95240; NID:g1262818; PIDN:CAA64527.1; PID:g1262819
 C:Genetics:
 A:Gene: SGP28
 C:Superfamily: cysteine-rich secretory protein 1
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:20-245/Product: neutrophil granules matrix glycoprotein SGP28 #status predicted <MAT>

Query Match 10.7%; Score 295.5; DB 2; Length 245;
 Best Local Similarity 31.0%; Pred. No. 1.9e-15;
 Matches 75; Conservative 29; Mismatches 91; Indels 47; Gaps 10;
 QY 13 LLEVLGSGGYLLPNVTLLELLSKYOHNSHVRRAIPREDKEILMLHKLKLG 69
 DB 7 LLEVLGSGGYLLPNVTLLELLSKYOHNSHVRRAIPREDKEILMLHKLKLG 52
 QY 70 VOPQASNMETWMDLELESAAMASQCIWEG-PTSLIVS-IGONTGAHWGRY 121
 DB 53 VSPFARNMLKMEWKNEMANNAQWMAQCNVRIKSNPKDRMTSLKGENLYMSSAPSSWSQ 111
 QY 122 RSPGFVQSWYDEKDYTPYSECPNCPERCSPGPMCTHYQIYVATNTKIGCAVNTIC 181
 DB 112 ----AIGSFEDYNDPFG---GVGPKPNAVVG---HYIQVWYSILVCGGNAYCP 158
 QY 182 KMTWGEWENAVYFCVNTSPKGNWIG--APYKNGRGPSECPSPYGGSCRNMLCYREE 239
 DB 159 NQKVL-----KYYVYQYCPAGNMANRLVYVEQGAPCASPDCDDGLCTGCKYEDL 212
 QY 240 YT 241
 DB 213 YS 214

RESULT 6
 B33329
 Cysteine-rich secretory protein 2 type I precursor - human
 N:Alternate names: testis-specific protein
 C:Species: Homo sapiens (man)
 C:Date: 09-Mar-1990 #sequence_revision 09-Mar-1990 #text_change 20-Jun-2000
 C:Accession: B33329; S68682
 R:Kaashara, M.; Gutknecht, J.; Brew, K.; Spurr, N.; Goodfellow, P.N.
 Genomics 5, 527-534, 1989
 A:Title: Cloning and mapping of a testis-specific gene with sequence similarity to a sperm
 A:Reference number: A33329; MUID:90129048
 A:Accession: B33329
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-243 <KAS>
 A:Cross-references: GB:M25532; NID:g339882; PIDN:AAA61220.1; PID:g339883
 R:Kraetzschmar, J.; Haendler, B.; Eberspaecher, U.; Roosterman, D.; Donner, P.; Schleunig, J.; Blochem, 236, 827-836, 1996

A:Title: The human cysteine-rich secretory protein (CRISP) family. Primary structure
 A:Reference number: S68681; MUID:96270732
 A:Accession: S68682
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-243 <KRA>
 A:Cross-references: EMBL:X95239; NID:g1262816; PIDN:CAA64526.1; PID:g1262817
 C:Genetics:
 A:Gene: GDB:TPX1
 A:Cross-references: GDB:120760; OMIM:187430
 A:Map position: 6p21-qter
 C:Superfamily: cysteine-rich secretory protein 1
 F:1-20/Domain: signal sequence #status predicted <SIG>
 F:21-243/Product: cysteine-rich secretory protein 2 type I #status predicted <MAT>

Query Match 10.2%; Score 281; DB 2; Length 243;
 Best Local Similarity 26.3%; Pred. No. 2.5e-14;
 Matches 83; Conservative 33; Mismatches 83; Indels 116; Gaps 14;
 QY 24 LLEVLGSGGYLLPNVTLLELLSKYOHNSHVRRAIPREDKEILMLHKLKLG 68
 DB 3 LLEVLGSGGYLLPNVTLLELLSKYOHNSHVRRAIPREDKEILMLHKLKLG 49
 QY 69 VOPQASNMETWMDLELESAAMASQCIWEG-PTSLIVS-IGONTGAHWGRYRSP 124
 DB 50 AVSPFARNMLKMEWKNEMANNAQWMAQCNVRIKSNPKDRMTSLKGENLYMSSAPSSWSQ 103
 QY 125 ----GFVQSWYDEKDYTPYSECPNCPERCSPGPMCTHYQIYVATNTKIGCAVNTIC 180
 DB 104 PTSSSAIQSWYDEKDYTPYSECPNCPERCSPGPMCTHYQIYVATNTKIGCAVNTIC 155
 QY 181 RMTWGEWENAVYFCVNTSPKGNWIG--EAPYKNGRGPSECPSPYGGSCRNMLCYREE 238
 DB 156 PNOD-----SLKYYVYQYCPAGNMANRLVYVEQGAPCASPDCDDGLCTGCKYEDL 201
 QY 239 TYTPKPEIDENNEVEFAPPEENHVLQPRVPRTPKRTSNAVYMTQVVRCDT 292
 DB 202 -----TMSQYODLLSNCDLKNFTAG 222
 QY 293 ---KMKDRKGSTC 303
 DB 223 CEHELKCK-ATC 236

RESULT 7
 T16415
 Hypothetical protein F48B.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
 C:Accession: T16415
 R:Kirsten, J.
 submitted to the EMBL Data Library, March 1995
 A:Description: The sequence of C. elegans cosmid F48B.
 A:Reference number: S59413
 A:Accession: T16415
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-312 <KIR>
 A:Cross-references: EMBL:U23514; NID:g746484; PID:g746485; PIDN:AAAC6538.1; CESP:F48E
 C:Experimental source: strain Bristol N2
 C:Genetics:
 A:Gene: CESP:F48E.1
 A:Introns: 28/1; 94/3; 153/3; 174/3; 226/3
 Query Match 9.7%; Score 269; DB 2; Length 312;
 Best Local Similarity 28.4%; Pred. No. 2.8e-13;
 Matches 83; Conservative 36; Mismatches 93; Indels 80; Gaps 17;
 QY 21 QGYLLPNVTLLELLSKYOHNSHVRRAIPREDKEILMLHKLKLG 70
 DB 55 RGYFPFHFQSDGLSRSEHPNDEYLRKWTTHNNRYRRNP----- 96

Query 71 QPQASNMETWDELLEKSAAMASOCIEHGPISLVSIGONICAHMGRYRSGEFHVS 130
 Db 97 --ASDMNMLYSDLELASAQRHADTCDFRHSRGR--INVENTWA--APYSNDASISI 149
 QY 131 WYDEKDYTPYPSRCNWCPC-----ERCSGPMCTHYTOIYATWATNKIGCVNTRCKMT- 184
 Db 150 WFNHVS-----NFRCCGNHAYKHC-----CGHYVQVWMAKTNLVCGRFCRVOG 195
 QY 185 WGEWENAVYVCNYSKGNMI-----GE-----APYKNGRCPSCPSYSGSCR 230
 Db 196 WVGCHGRNV--FVCHYNDQGNTEVYTAGQIYAMPATWASGDGK--CSNCPAN--APACY 251
 QY 231 NNLCYREBTYPKPEDEMEVETAPICEENHWLQPRVAKPTKPKTSAYN 282
 Db 252 QGLCYMKKNYEAPTTTE--STTSTTTE-----PTTCPEPEPEAGADN 296

RESULT 8
 JE0204
 C:Species: Rattus norvegicus (Norway rat)
 C>Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 21-Jul-2000
 C:Accession: JE0204
 R:Maeda, T.; Sakashita, M.; Ohba, Y.; Nakanishi, Y.
 Biochem. Biophys. Res. Commun. 248, 140-146, 1998
 A>Title: Molecular cloning of the rat tpx-1 responsible for the interaction between spect
 A:Reference number: JE0204; MUID:98340864
 A:Accession: JE0204
 A:Molecule type: mRNA
 A:Residues: 1-243 <MAE>
 A:Cross-references: DDBJ:AB009662; NID:93374579; PIDN:RAA22029.1; PID:93374580
 C:Comment: This protein functions as a cell adhesion protein for the association between
 C:Genetics:
 A:Map position: 17
 C:Superfamily: cysteine-rich secretory protein 1

Query Match 9.7%; Score 268.5; DB 2; Length 243;
 Best Local Similarity 31.7%; Pred. No. 2,3e-13;
 Matches 60; Conservative 29; Mismatches 65; Indels 35; Gaps 8;

QY 56 KEELMLNKLKRGVQPOQASNMETWDELLEKSAAMASOCIEHGPISLVSIGQ 112
 Db 38 QREITAKNELRROVSPGNSILKEMVQAANAOKWANNICIEHSTEDRKINICGE 97
 QY 113 NL-----GAHMGYRSPGFHVSQWYDEKDYTPYPSRCNWCPCERSGPMCTHYTOIW 167
 Db 98 NLVYSTDPTSMRTV-----IQSWYEEENENFVGVGAK-----PNSAVG-----HYTOLW 142
 QY 168 ATTAKICAVNTCKKMTWGEWENAVYVCNYSKGNMI--GEAPYKNGRCPSCPSY 225
 Db 143 YSSKVKCGVAVCPNODTL-----KYFYVCHYCPMGNNVNMKSTYHQSTPCASCP--- 193
 QY 226 GSGCRNNLC 234
 Db 194 -NNCDNGLC 201

RESULT 9
 A33329
 C:Species: Mus musculus (house mouse)
 C>Date: 09-Mar-1990 #sequence_revision 09-Mar-1990 #text_change 29-Sep-1999
 C:Accession: A33329
 R:Kasahara, M.; Gutknecht, J.; Brew, K.; Spurr, N.; Goodfellow, P.N.
 Genomics 5, 527-534, 1989
 A>Title: Cloning and mapping of a testis-specific gene with sequence similarity to a spe
 A:Reference number: A33329; MUID:90129048
 C:Accession: A33329
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-243 <KAS>

A:Cross-references: GB:M25533; NID:920126; PIDN:AAA0472.1; PID:9202127
 C:Superfamily: cysteine-rich secretory protein 1

Query Match 9.7%; Score 267.5; DB 2; Length 243;
 Best Local Similarity 33.2%; Pred. No. 2.8e-13;
 Matches 63; Conservative 28; Mismatches 62; Indels 37; Gaps 9;

QY 56 KEELMLNKLKRGVQPOQASNMETWDELLEKSAAMASOCIEHGPISLVSIGQ 112
 Db 38 QREITAKNELRROVSPGNSILKEMVQAANAOKWANNICIEHSTEDRKINICGE 97
 QY 113 NL-----GAHMGYRSP-----GFHVSQWYDEKDYTPYPSRCNWCPCERSGPMCTHYTOIW 167
 Db 98 NL-----YMSIDPTLMSVTSIOWSWINENDEPYGVGAK-----PNSAVG-----HYTOLW 142
 QY 168 ATTAKICAVNTCKKMTWGEWENAVYVCNYSKGNMI--GEAPYKNGRCPSCPSY 224
 Db 143 YSSKVKCGIAVCPNQ-----DNLKIFYVCHYCPMGNNVNMKSTYHQSTPCASCP-- 193
 QY 225 YGSGCRNNLC 234
 Db 194 -NNCENGIC 201

RESULT 10
 A44583
 C:Species: Vespa flavopilosa
 C>Date: 27-Jun-1994 #sequence_revision 27-Jun-1994 #text_change 11-Jan-2000
 C:Accession: A44583; B44522
 R:Hoffman, D.R.
 J. Allergy Clin. Immunol. 92, 707-716, 1993
 A>Title: Allergens in hymenoptera venom XXV: the amino acid sequences of antigen 5 mo
 A:Reference number: A44583; MUID:9404316
 A:Accession: A44583
 A>Status: preliminary
 A:Molecule type: protein
 A:Residues: 1-204 <HOP>
 C:Superfamily: yellowjacket venom allergen antigen 5

Query Match 9.2%; Score 253; DB 2; Length 204;
 Best Local Similarity 30.5%; Pred. No. 3e-12;
 Matches 71; Conservative 32; Mismatches 74; Indels 56; Gaps 12;

QY 1 MSCVLGVPLGLFLVCGSGGILLPNTLLLELLSKYOHNSHSRVRAIPREDKEIL 60
 Db 6 IKCLAGV-----HTAC--KYSLKPNCG--NKVVVS-----GLTKQEKDIL 45
 QY 61 MLHNKLKRGV-----QPOASNMETWDELLEKSAAMASOCIEHGPISLVS 108
 Db 46 KEHNDPQKILANGLETGRNGPQPAKKNKLNWMDLAVYAQVWANOCCYGHCTCRDIA 105
 QY 109 --SIGONL-----GAHMGYRSPGFHVSQWYDEKDYTPYPSRCNWCPCERSGP---MCT 160
 Db 106 KYQVGNVALGSTAKYDDPYKILKWMEDVKDY-----NP--KKKSGNNFLKTG 155
 QY 161 HTQIYMAATNKIGCAVNTCKKMTWGEWENAVYVCNYSKGNMI--GEAPYKNGRCPSCPSY 213
 Db 156 HTQVWMAVNTKEVCG-----SIFIOEKMKH--YLVCNCGPSGNFONEELYQ 202

RESULT 11
 B37350
 C:Species: Solenopsis invicta (red imported fire ant)
 C>Date: 01-Oct-1992 #sequence_revision 01-Oct-1992 #text_change 11-Jan-2000
 C:Accession: C44582; B60727; B37350
 R:Hoffman, D.R.
 J. Allergy Clin. Immunol. 91, 71-78, 1993
 A>Title: Allergens in hymenoptera venom XXIV: the amino acid sequences of imported fl

RESULT 13
B37329

THE UNIVERSITY OF CHICAGO

Oy 178 NCRKMTVNGEVMENAVYFVCNYS PKGN--WIGELPYKNGRPCSECPDPSYGSSCRNNLC- 23
::: ::::

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 25, 2002, 14:28:55 ; Search time 63.23 seconds
(without alignments)
304.343 Million cell updates/sec

Title: us-09-667-380a-2

Perfect score: 2759
Sequence: 1 MGVGLGVPLGLFLVCGS.....SESLGTPDQKAFRLFAVRQ 497

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	355	12.9	266	1 GLIP_HUMAN	P48060 homo sapien
2	301.5	10.9	246	1 AEG_RAT	P12020 ratius norv
3	298	10.8	244	1 AEG1_MOUSE	Q03401 mus musculu
4	298	10.8	245	1 CRG3_HORSE	O19010 equus cabal
5	297.5	10.8	245	1 CRG3_HUMAN	P54108 homo sapien
6	281	10.2	243	1 TPX1_HUMAN	P16562 homo sapien
7	269	9.7	312	1 YR81_CAEEL	Q09566 caenorhadi
8	268.5	9.7	244	1 TPX1_MOUSE	Q60477 cavia porce
9	267.5	9.7	243	1 TPX1_MOUSE	P16563 mus musculu
10	257.5	9.3	242	1 HELO_HELHO	O01055 heloderma h
11	256	9.2	227	1 VAS_VESPU	O05110 vesputia vul
12	253	9.2	204	1 VAS_VESPU	P35783 vesputia fla
13	244	8.8	204	1 VAS_VESPU	P35785 vesputia pen
14	243.5	8.8	234	1 VAS_SOLIN	P35778 solenopsis
15	240	8.7	204	1 VAS_VESMC	P35760 vesputia mac
16	239.5	8.7	249	1 CRG1_HUMAN	P54107 homo sapien
17	238	8.6	202	1 VAS1_VESCR	P35784 vesputia ger
18	238	8.6	204	1 VAS_VESGE	P35787 vesputia vid
19	238	8.6	206	1 VAS_VESVI	P35782 vesputia crabr
20	237	8.6	202	1 VAS2_VESCR	O05108 dolichovesp
21	237	8.6	203	1 VAS_DOLAR	P35779 solenopsis
22	237	8.6	211	1 VAS_SOLAR	P35777 ancylostoma
23	236.5	8.6	424	1 ASP_ANCA	P81657 vespa manda
24	235	8.5	202	1 VAS_VESMA	P35786 vesputia squ
25	235	8.5	205	1 VAS_VESPO	P35780 polistes fu
26	233.5	8.4	205	1 VAS_POLFU	O40374 medicago tr
27	232	8.3	206	1 VAS_POIDO	P81656 polistes do
28	230	8.1	205	1 VAS_POLEX	P35759 polistes ex
29	222.5	8.1	209	1 VAS3_POLAN	O05109 polistes an
30	222.5	8.0	215	1 VAS3_DOLMA	P10737 dolichovesp
31	221.5	8.0	183	1 CRVP_TRIMU	P79845 trimeresu
32	219.5	7.9	227	1 VAS2_DOLMA	P10736 dolichovesp
33	217				

34	213.5	7.7	204	1 SC7_SCHCO	P35794 schizophyl1
35	212	7.7	241	1 AEG2_MOUSE	Q03402 mus musculu
36	208.5	7.6	168	1 PRLA_TOBAC	P08299 nicotiana t
37	207	7.5	177	1 PRL1_TOBAC	P11670 nicotiana t
38	206.5	7.5	159	1 PR04_LYCES	O04108 lycopersico
39	203.5	7.4	167	1 PRMS_MAIZE	Q00008 zea mays (m
40	202.5	7.3	168	1 PRL1_TOBAC	P07053 nicotiana t
41	200.5	7.3	168	1 PRL1_TOBAC	P09042 nicotiana t
42	199.5	7.2	214	1 SC14_SCHCO	P35795 schizophyl1
43	199	7.2	164	1 PRL1_HORVU	P35793 hordeum vull
44	198	7.2	164	1 PRL1_HORVU	P35793 hordeum vull
45	198	7.2	164	1 PRL1_HORVU	Q05968 hordeum vull

ALIGNMENTS

RESULT 1
GLIP_HUMAN STANDARD; PRT; 266 AA.

AC P48060: 015409;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glioma pathogenesis-related protein (Glipr) (RTVP-1 protein).
GN GLIPR OR RTVP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP TISSUE=glial tumor;
RC MEDLINE=97128816; PubMed=8973356;
RA Rich T., Chen P., Furman F., Huynh N., Israel M.A.;
RT RTVP-1, a novel human gene with sequence similarity to genes of
RT diverse species, is expressed in tumor cell lines of glial but not
RT neuronal origin. (1996).
RL Gene 180:125-130(1996).
RN [2]
RP SEQUENCE OF 11-266 FROM N.A.
RX MEDLINE=9531646; PubMed=7607567;
RA Murphy E.V., Zhang Y., Zhu W., Biggs J.;
RT "The human glioma pathogenesis-related protein is structurally
RT related to plant pathogenesis-related proteins and its gene is
RT expressed specifically in brain tumors.";
RL Gene 159:131-135(1995).
RN [3]
RP STRUCTURE BY NMR.
RX MEDLINE=98151500; PubMed=9482873;
RA Szyperki T., Fernandez C., Mumenthaler C., Wuehrlich K.;
RT "Structure comparison of human glioma pathogenesis-related protein
RT Glipr and the plant pathogenesis-related protein P1a indicates a
RT functional link between the human immune system and a plant defense
RT system.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:2262-2266(1998).

- I- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN THE HUMAN BRAIN TUMOR,
GLOBULASTOMA MULTIFORM/ASTROCYTOMA, BUT NEITHER IN NORMAL FETAL
OR ADULT BRAIN TISSUE, NOR IN OTHER NERVOUS SYSTEM TUMORS.
- I- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MAMMALIAN SCP/TPX1;
INSECTS AG3/AG5; FUNGI SC7/SC14 AND PLANTS PR-1.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL: X91911; CAA63005.1;
DR EMBL: U16307; AAA82731.1;
HSP: P04284; ICFE.

DR MIM: 602692; -
 DR Interp: IPR001283; SCP.
 DR Pfam: PF00188; SCP; 1.
 DR PRINTS: PR00837; V5FPLIKE.
 DR SMART: SM00198; SCP; 1.
 DR PROSITE: PS01009; SCP_AGS_PRL_SC7_1; 1.
 DR PROSITE: PS01010; SCP_AGS_PRL_SC7_2; 1.
 FT CONFLICT 125 D -> N (IN REF. 2).
 FT CONFLICT 209 266 RORDQKRYSVYVPGWIPRNPRTSLFLVNSVILLV
 ITTIVOLKYPVLVLD -> DSEKSNVTTMLTIRLAHIS
 T (IN REF. 2)
 SQ SEQUENCE 266 AA: 30342 MW: C04378085FEB2B7A CRC64;

Query Match
 Best Local Similarity 12.9%; Score 355; DB 1; Length 266;
 Matches 82; Conservative 33; Mismatches 75; Indels 40; Gaps 12;

QY 25 LPNVLLEELSKYOHNSHVRRAIPRED-KEEILMLHKLKRGQVQPOASNEEYMTWD 83
 DB 5 LATIAMVSVFVSNYSHF---ANILPDIENEDPDKCVRIHKKFSEVKEPTASDMLYMTWD 61
 QY 84 DELESAAMAASOCIWEHG---PTSL---LVSIGMLGAMGRYSPGPHYOS---W 131
 DB 62 PALAOIAKAMASNCOFSHNTRLRKPPHKLHPFTSLGENI---W-TGSVPLITSALTMTW 117
 QY 132 YDEKVDYTPYSPCNPCPCRCGPMCTHTYTQVATYTKIGCAVNTCRKMTVGEWE 191
 DB 118 YDEIQDIDF-----KTRICKKVCGHYTOVMADSKYVCAVOFCPRKSGF-DALS 166
 QY 192 NAVFVVCNYSFSGN---WISGAPYKNGRPGCEPSPSYGSGCRNNLCYREE 238
 DB 167 NGAFHICNYGPGGNYPTW---PYKRGATCSACPNN--DKCLDNLCVNRQ 210

RESULT 2
 AEG_RAT STANDARD: PRT; 246 AA.

AC P12020;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, last annotation update)
 DE Sperm-coating glycoprotein precursor (SCP) (Acidic epididymal glycoprotein) (Protein D) (Protein E) (Protein IV) (Sialoprotein)
 DE (32 kDa epididymal protein).
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 NCBI_Taxid=10116;
 RP SEQUENCE FROM N.A.
 RC TISSUE-Epididymis;
 RC MEDLINE=87053955; PubMed=3780731;
 RA Brooks D.E., Means A.R., Wright E.J., Singh S.P., Tiver K.K.;
 RA "Molecular cloning of the cDNA for androgen-dependent sperm-coating glycoproteins secreted by the rat epididymis";
 RT Eur. J. Biochem. 161:13-18(1986).
 RN [12]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89039913; PubMed=2460753;
 RA Charrest N.J., Joseph D.R., Wilson E.M., French F.S.;
 RA "Molecular cloning of complementary deoxyribonucleic acid for an androgen-regulated epididymal protein: sequence homology with metalloproteins";
 RT Mol. Endocrinol. 2:999-1004(1988).
 RL [1]
 CC -1- FUNCTION: THIS PROTEIN IS SUPPOSED TO HELP SPERMATOZOA UNDERGO FUNCTIONAL MATURATION WHILE THEY MOVE FROM THE TESTIS TO THE DUCTUS DEFERENS.
 CC -1- INDUCTION: By androgens.
 CC -1- MISCELLANEOUS: SCP IS AN ANDROGEN-DEPENDENT PROTEIN, WHICH IS SECRETED BY THE EPIDIDYMAL EPITHELIUM AND THEN BECOMES ASSOCIATED WITH THE SPERM SURFACE. TWO MAJOR VARIANT PROTEIN D AND E DIFFER FROM EACH OTHER BY THEIR CARBOHYDRATE SIDE CHAINS.

CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MAMMALIAN SCP/TPX1;
 CC INSECTS AG3/AG5; FUNGI SC7/SC14 AND PLANTS PR-1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: X04643; CA28304.1; -
 DR EMBL: M31173; BAB59716.1; -
 DR PIR: A24609; A24609.
 DR PIR: A40918; A40918.
 DR Interp: IPR001283; SCP.
 DR Pfam: PF00188; SCP; 1.
 DR PRINTS: PR00837; V5FPLIKE.
 DR SMART: SM00198; SCP; 1.
 DR PROSITE: PS01009; SCP_AGS_PRL_SC7_1; 1.
 DR PROSITE: PS01010; SCP_AGS_PRL_SC7_2; 1.
 KW Sperm; Glycoprotein; Signal.
 FT SIGNAL 1 19
 FT CHAIN 20 246
 FT MOD_RES 20 246
 FT CARBOHYD 32 32
 FT CARBOHYD 85 85
 FT CARBOHYD 147 147
 FT CARBOHYD 213 213
 SQ SEQUENCE 246 AA: 27847 MW: 585207C7CF7CE9D1 CRC64;

Query Match
 Best Local Similarity 10.9%; Score 301.5; DB 1; Length 246;
 Matches 81; Conservative 51; Mismatches 90; Indels 81; Gaps 15;

QY 11 LGILFLVCGSGYGLPNVTLLEELSKYOHN-ESHSRRRAIPREDKEEILMLHKLKRGQ 69
 DB 5 LVLEFLAA-----VLP-PSLIQDTTDEMDRIENISTTKLSV---OEELINKHQLRRT 54
 QY 70 VQPOASNEEYMTWDELEKSAAMAASOCIWEHG---TSLVLSIGONL-----CAHGRY 121
 DB 55 VSPSSDILREVMHDVAVNAQKMANRCIYNHSPLOHRTTLKKGEMLFMANIPASVSV 114
 QY 122 RSPGFHVQSWYDEKDYTPYSPCNPCPCRCGPMCTHTYTQVATYTKIGCAVNTCR 181
 DB 115 -----IQDWDESLDVPFGG-----PKV-GVAVGHYTOVMNSVTLVACVAECP 160
 QY 182 KATYVGEWENAVFVCNYSFSGNMIGE--APYKNGRPGCEPSPSYGSGCRNNLCYREE 239
 DB 161 DQPL-----KYFYVCHYCPGCGNYVGRILSPYTEGRDSCF-----GNCEDGLCTNSCE 209
 QY 240 YTPRPETDENNEVETAPIPENNVHVLQPRMRTPKRKTSAAVVMQVYVRCDTK-KKRC 298
 DB 210 Y-----EDNT-----SNCGLKKNVSCDDPLKKEC 235
 QY 299 KGS 301
 DB 236 RAS 238

RESULT 3
 AEG1_MOUSE

AC 003401;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, last annotation update)
 DE Sperm-coating glycoprotein I precursor (SCP 1) (Acidic epididymal glycoprotein 1) (Cysteine-rich secretory protein-1) (Crisp-1).
 NCBI_Taxid=10116;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Submandibular gland;
 RX MEDLINE=93246016; PubMed=1301383;
 RA Mizuki N., Kasahara M.;
 RT "Mouse submandibular glands express an androgen-regulated transcript
 RT encoding an acidic epididymal glycoprotein-like molecule.";
 RT Mol. Cell. Endocrinol. 89:25-32(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Submandibular gland;
 RX MEDLINE=93307144; PubMed=8319566;
 RA Haendler B., Kratzschmar J., Theuring F., Schleuning W.D.;
 RT "Transcripts for cysteine-rich secretory protein-1 (CRISP-1; DE/ABC)
 RT and the novel related CRISP-3 are expressed under androgen control in
 RT the mouse salivary gland";
 RT Endocrinology 133:192-198(1993).
 CC -1 FUNCTION: THIS PROTEIN IS SUPPOSED TO HELP SPERMATOZOA UNDERGO
 CC FUNCTIONAL MATURATION WHILE THEY MOVE FROM THE TESTIS TO THE
 CC DUCTUS DEFERENS.
 CC -1 SUBCELLULAR LOCATION: STORED IN SECRETORY GRANULES OF GRANULAR
 CC CONVOLUTED TUBULES CELLS.
 CC -1 TISSUE SPECIFICITY: MAINLY FOUND IN THE CAUDA EPIDIDYMIS WHERE IT
 CC IS SYNTHESIZED BY THE PRINCIPAL CELLS AND SECRETED INTO THE LUMEN.
 CC BINS TO THE HEADS OF SPERMATOZOA. ALSO EXPRESSED IN THE
 CC SUBMANDIBULAR GLAND.
 CC -1 DEVELOPMENTAL STAGE: EXPONENTIAL INCREASE BETWEEN DAYS 25 AND 30
 CC AFTER BIRTH.
 CC -1 INDUCTION: By androgens.
 CC -1 SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MAMMALIAN SCP/TPX1;
 CC INSECTS AG3/AG5; FUNGI SC7/SC14 AND PLANTS PR-1.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M92849; AAA37185.1; -
 DR EMBL: M95559; AAA37460.1; -
 DR PIR: A49202; A49202.
 DR MGI: MGI:102553; Aeg1.
 DR InterPro: IPR001283; SCP.
 DR Pfam: PF00188; SCP; 1.
 DR PRINTS: PRO0837; V5TPXLIKE.
 DR SMART: SM00198; SCP; 1.
 DR PROSITE: PS01009; SCP_AGS_PRL_SC7_1; 1.
 DR PROSITE: PS01010; SCP_AGS_PRL_SC7_2; 1.
 DR Signal: Glycoprotein; Signal.
 FT SIGNAL 1 19
 FT CHAIN 20 244
 FT CARBOHYD 145 145 N-LINKED (GLCNAC...) (POTENTIAL).
 FT SEQUENCE 244 AA; 27679 MW; D0DD0348F85781F CRC64;
 SO
 Query Match 10.8%; Score 298; DB 1; Length 244;
 Best Local Similarity 32.0%; Pred. No. 1,le-16;
 Matches 77; Conservative 35; Mismatches 87; Indels 42; Gaps 11;
 OY 11 IGLFLVCGSGGYLTPNVLLELLSKYQHNESHVRAIRAREDEKEELIMHNKIRGOV 70
 DB 5 LVLFELAA-----VLP-PSILOD-SSQENLEKLSTFKMSV-----OEELVSKHNOIRRV 53
 OY 71 QPOASNMETWTWDELEKSAAMAASOCIMWEGHPTSLVS-----IGONIG-----AHGGRYR 122
 DB 54 SPSSGSDILKEMNYDAQVNAQOMADCTSHSIELRTTLRCGHNLFMSYLAASS-- 111
 OY 123 SPGHVOSWYDEVKDYTPYPSECNPCPCRCGPMCTHTOTATVATNKIGCAVNTCKR 182
 DB 123 SPGHVOSWYDEVKDYTPYPSECNPCPCRCGPMCTHTOTATVATNKIGCAVNTCKR 182

DB 112 -----AIOGWNEKLDLY-----DVGPKQPSVYG-----HYQVWNSTFOVAGNAECPK 159
 OY 183 MTWGEWENAVYFVCCNYSPKGNWIGF--APYKNRGRPCSECPYSGSCNNLCYREETY 240
 DB 160 NPL-----RYYYVCHCPVGNYGRLYTPYRAGEFACSCPHCEDGLCTNCGHEDEK 212
 OY 241 T 241
 DB 213 T 213
 RESULT 4
 CR33 HORSE STANDARD; PRT; 245 AA.
 ID CR33 HORSE 019010;
 AC 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Cysteine-rich secretory protein-3 precursor (CRISP-3).
 GN CRISP3.
 OS Equus caballus (Horse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
 OX NCBI_TaxID=9796;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Amputilla;
 RX MEDLINE=98422318; PubMed=9748582;
 RA Schombony A., Gentzel M., Wolfes H., Raida M., Neumann U.,
 RA Toepfer-Petersen E.;
 RT "Equine CRISP-3: primary structure and expression in the male genital
 RT tract";
 RT Biochim. Biophys. Acta 1387:206-216(1998).
 CC -1 SUBCELLULAR LOCATION: SECRETED; IN NEUTROPHILS. LOCALIZED IN
 CC SPECIFIC GRANULES (BY SIMILARITY).
 CC -1 TISSUE SPECIFICITY: EXPRESSED IN THE SALIVARY GLAND, IN THE
 CC AMPULLA AND THE SEMINAL VESICLE.
 CC -1 SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MAMMALIAN SCP/TPX1;
 CC INSECTS AG3/AG5; FUNGI SC7/SC14 AND PLANTS PR-1.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AJ001400; CA04729.1; -
 DR HSSP: P04284; ICFE.
 DR InterPro: IPR001283; SCP.
 DR Pfam: PF00188; SCP; 1.
 DR PRINTS: PRO0837; V5TPXLIKE.
 DR SMART: SM00198; SCP; 1.
 DR PROSITE: PS01009; SCP_AGS_PRL_SC7_1; 1.
 DR PROSITE: PS01010; SCP_AGS_PRL_SC7_2; 1.
 KW Signal: Multigene family.
 FT SIGNAL 1 22
 FT CHAIN 23 245
 FT SEQUENCE 245 AA; 27308 MW; 8934AE87F402BA22 CRC64;
 SO
 Query Match 10.8%; Score 298; DB 1; Length 245;
 Best Local Similarity 27.9%; Pred. No. 1,le-16;
 Matches 77; Conservative 35; Mismatches 72; Indels 92; Gaps 13;
 OY 54 EKEKEELIMHNKIRGOVQPOASNMETWTWDELEKSAAMAASOCIMWEGHPTSLVS-----LLVS 109
 DB 36 EVQKEIVNKHNDLRRTVSPFLASNMILKQMDSTATNAQNMANKLLOSKAEDRAVGTMK 95
 OY 110 IGQNT-----GAHWGRYRSPGFHVOSWYDEVKDYTPYPSECNPCPCRCGPMCTHTO 164
 DB 96 CGENLFMSIIPMSWD-----AIONWHDVDFY-----GVGPTPAVAVG-----HTQ 141


```

CC      EMBL; M25532; AAA61220.1; -
DR      EMBL; X95239; CAA64526.1; -
DR      PIR; B33329; B33329.
DR      MIM; 187430; -
DR      InterPro: IPR001283; SCP.
DR      Pfam: PF00188; V5TPX1KE.
DR      PRINTS; SM00198; SCP. 1.
DR      SMART; SM00198; SCP. 1.
DR      PROSITE; PS01009; SCP_AGS_PRL_SCT_1; 1.
DR      PROSITE; PS01010; SCP_AGS_PRL_SCT_2; 1.
DR      Testis; Signal; Multigene family.
KW      SIGNAL 1 21 POTENTIAL.
FT      CHAIN 22 243 TESTIS-SPECIFIC PROTEIN TPR-1.
SQ      SEQUENCE 243 AA; 27259 MW; C5FE698C449CFA09 CRC64;

Query Match          10.2%; Score 281; DB 1; Length 243;
Best Local Similarity 26.3%; Pred. No. 2.4e-15;
Matches 83; Conservative 33; Mismatches 83; Indels 116; Gaps 14;

QY 24 LIPNVTLLEELISKYOHNSHVRRAIPREDKE-----ETLMHNLKRG 68
DB 3 LIPVPLVLYLPL-----SLPAEGKDPATLTLTLOLOVOREIYNKINELR 49
QY 69 QVOPQASNMETWTWDELEKSAANASOCIEHNG-PTSLVLS--IGONLGAHMGRRSP- 124
DB 50 AVSPASNMMLKMEHSREVTTNAORMAKCTLOHSDPEDKRTSRGENL-----YMSD 103
QY 125 -----GFVQSVYDEKDYTPYPSECPNCPGSCPMCTHTYQIYMATNKGCAVNTC 180
DB 104 PLSMSAIGSYDEILDFEY-----GVGPKSPNAVVG---HYTLVWYSTYQVGCIAVC 155
QY 181 RKMTVWGEVWENAVYFCNCSPPKGNWIG--EAPKNGRCPSECPSPSYGSCRNLCYREE 238
DB 156 PNOD-----SLKYYVCQYCPAGNNMKNKRTFYQOSTPCAGCP-----DDCDGLC- 201
QY 229 TTPKPEIDEMNEVETAPIPEENHWLQPRVMPPTKRTSAVYMYQVRYCPT----- 292
DB 202 -----TNSCOYQDLSNCDLSKNTAG 222

QY 293 -----KKKDRCKGSC 303
DB 223 CEHELKECK-ATC 236

RESULT 7
ID      YR81.CAEEL STANDARD; PRT; 312 AA.
AC      Q09566;
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Hypothetical 35.0 kDa protein F48B8.1 in chromosome III.
GN      F48B8.1.
OS      Caenorhabditis elegans.
OC      Caenorhabditidae; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC      Eukaryota; Metazoa; Nematoda; Rhabditidae; Rhabditidae;
OC      Rhabditidae; Peleodermidae; Caenorhabditidae.
OX      NCBI_TaxID=6239;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=BRISTOL N2;
RL      Kistner J.;
RL      Submitted (MAR-1995) to the EMBL/GenBank/DDBJ databases.
CC      -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MAMMALIAN SCP/TPX1;
CC      INSECTS AG3/AG5; FUNGI SC7/SC14 AND PLANTS PR-1.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb.ch/announce/

```

```

CC      or send an email to license@isb-sib.ch).
CC      EMBL; U23514; AAC46538.1; -
DR      HSSP; P04284; ICFE.
DR      WormPep; F48B8.1; CE01953.
DR      InterPro: IPR001283; SCP.
DR      Pfam; PF00188; SCP. 1.
DR      PRINTS; PR00837; V5TPX1KE.
DR      SMART; SM00198; SCP. 1.
DR      PROSITE; PS01009; SCP_AGS_PRL_SCT_1; FALSE NEG.
DR      PROSITE; PS01010; SCP_AGS_PRL_SCT_2; FALSE NEG.
DR      Hypothetical protein.
KW      DOMAIN 265 268 POLY-THR.
FT      DOMAIN 299 303 POLY-GLU.
SQ      SEQUENCE 312 AA; 35054 MW; AEFCTBF25E26288 CRC64;

Query Match          9.7%; Score 269; DB 1; Length 312;
Best Local Similarity 28.4%; Pred. No. 2.9e-14;
Matches 83; Conservative 36; Mismatches 93; Indels 80; Gaps 17;

QY 21 QGILLNVTLLEL-LISKYOH-NE-----SHSRVRAIPREDKEETLMHNLKRGV 70
DB 55 RGYFFPSHFQSDGLSRSEHPEYIKKWTIEHNRRYRMP----- 96
QY 71 QPQASNMETWTWDELEKSAANASOCIEHNGPTSLVLSIGONLGAHMGRRSPGFHYQS 130
DB 97 -----ASDNMMLYMSDELAAASORHADTCDFHRSGR--INVENINIA--APYSNDAISI 149
QY 131 WYDEKDYTPYPSECPNCP-----ERCSGPMCTHYTOIYMATNKGCAVNTCRKRT- 184
DB 150 WFNELV-----NPRGCHNAHYKC-----CGHYGVYVMAKTLVCGGSPRCRDVVG 195
QY 185 VMGEVWENAVYFCNCSPPKGNWIT-----GE-----APYNGRCPSECPSPSYGSGSCR 230
DB 196 VMGRCHRNW--FVCHYXNOGNTVFTARGOLYAPAFMTASGNGK-CNSCPAN-APACY 251
QY 231 NNLCTREETYPKPEIDEMNEVETAPIPEENHWLQPRVMPPTKRTSAVY 282
DB 252 QGLCYMPKMYEAPITTYTB--STTSTTTEE-----PTTCEPDEPEEAGADN 296

RESULT 8
ID      TPX1.CAVPO STANDARD; PRT; 244 AA.
AC      Q60477;
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      01-NOV-1997 (Rel. 35, Last annotation update)
DE      Testis-specific protein TPX-1 precursor (Autoantigen 1) (25 kDa
DE      acrosomal autoantigen) (Aa1).
GN      TPX1.
OS      Cavia porcellus (Guinea pig).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Euteria; Rodentia; Hystriocognathi; Cavidae; Cavia.
OX      NCBI_TaxID=10141;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=HARTLEY; TISSUE=Testis;
RC      MEDLINE=96354287; PubMed=9115720;
RA      Foster J.A., Gerton G.L.;
RT      "Autoantigen 1 of the guinea pig sperm acrosome is the homologue of
RT      mouse tpX-1 and human TPX1 family-";
RT      secretory protein (CRISP) family-";
RL      Mol. Reprod. Dev. 44:221-229(1996).
RN      [2]
RP      SEQUENCE OF 22-41. PubMed=3282555;
RA      Hardy D.M., Huang T.T.F., Jr., Driscoll W.J., Tung K.S.K., Wild G.C.;
RT      "Purification and characterization of the primary acrosomal
RT      autoantigen of guinea pig epididymal spermatozoa.";
RL      Biol. Reprod. 38:423-437(1988).
CC      -1- TISSUE SPECIFICITY: TESTIS.

```

```

CC -! SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MAMMALIAN SCP/TPX1;
CC INSECTS AG3/AG5; FUNGI SC7/SC14 AND PLANTS PR-1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; U35712; AAC52616.1; -
DR HSSP; P04284; 1CFE.
DR InterPro; IPR001283; SCP.
DR Pfam; PF00188; SCP; 1.
DR PRINTS; PR00837; V5TPXLIKE.
DR SMART; SM00198; SCP; 1.
DR PROSITE; PS01009; SCP_AGS_PRL_SC7_1; 1.
DR PROSITE; PS01010; SCP_AGS_PRL_SC7_2; 1.
DR KMW Testis; Signal.
FT SIGNAL 1 21
FT CHAIN 1 21
SQ SEQUENCE 244 AA; 27248 MW; 58DBDE6ECEE16A12 CRC64;

Query Match
Best Local Similarity 9.7%; Score 268.5; DB 1; Length 244;
Matches 74; Conservative 29; Mismatches 87; Indels 67; Gaps 12;

QY 56 KEELIMLNKLRGOVQOASNMETWDELEKSAAMASQCIEMHG-----PTSLIVSI 110
DB 38 QREIYNKHNELRSVNPFSIDILKEMSIQATINAKANKCIIHSSKDKRKINIGCE 97
QY 111 GQNL-----GAHGRYRSPGFHVSQWYDEKYDTYTPY-PSECNFWCPCRGSCPMCTHYTQ 164
DB 96 GENLYMSDPSWSD-----AIQSFEDSODFTFEVGRKSHN-----AVGCHYTQ 140
QY 165 IYVATNKIKGAVNCRKKTWGEVWENAVYVCNYSKPGN--WIGEAPYKNGRSCSECP 222
DB 141 LVMYSYLVGCGIAYCPNOD-----SLKYYVYCYCPAGNNVYIKNPYKQIGTCASCP 194
QY 223 PEGGSCRNMLCYREBYTPKPEFDEMNEVEETAPIPENHNWLOPRVWPKPKKTSAVN 282
DB 195 ---CHCENGICLTNCEY-----EDLLNCESL-----KNTACE 225
QY 283 YMTQVVRGDTKMKDKCK 299
DB 226 HQLLYEKC--KATCRCE 240

RESULT 9
TPX1_MOUSE
ID TPX1_MOUSE STANDARD; PRT: 243 AA.
AC P16563;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Testis-specific protein TPX-1 precursor.
GN TPX1 OR TPX-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=90129048; PubMed=2613236;
RA Kasahara M., Gutknecht D., Brew K., Spurr N., Goodfellow P.N.;
RT "Cloning and mapping of a testis-specific gene with sequence
RT similarity to a sperm-coating glycoprotein gene.";
RL Genomics 5:537-534(1989).
CC -! SUBCELLULAR LOCATION: Secreted (Probable).
CC -! TISSUE SPECIFICITY: TESTIS.

```

```

CC -! SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MAMMALIAN SCP/TPX1;
CC INSECTS AG3/AG5; FUNGI SC7/SC14 AND PLANTS PR-1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; M25533; AAA40472.1; -
DR PIR; A33329; A33329.
DR HSSP; P04284; 1CFE.
DR MGD; MGI:98815; TPX1.
DR InterPro; IPR001283; SCP.
DR Pfam; PF00188; SCP; 1.
DR PRINTS; PR00837; V5TPXLIKE.
DR SMART; SM00198; SCP; 1.
DR PROSITE; PS01009; SCP_AGS_PRL_SC7_1; 1.
DR PROSITE; PS01010; SCP_AGS_PRL_SC7_2; 1.
DR KMW Testis; Signal.
FT SIGNAL 1 22
FT CHAIN 1 22
SQ SEQUENCE 243 AA; 27605 MW; 6E707E569ACAA244 CRC64;

Query Match
Best Local Similarity 9.7%; Score 267.5; DB 1; Length 243;
Matches 63; Conservative 28; Mismatches 62; Indels 37; Gaps 9;

QY 56 KEELIMLNKLRGOVQOASNMETWDELEKSAAMASQCIEMHG-----PTSLIVSIQ 112
DB 38 QREIYNKHNELRSVNPFSIDILKEMSIQATINAKANKCIIHSSKDKRKINIGCE 97
QY 113 NLGAHWGRYRSP-----GFHVSQWYDEKYDTYTPY-PSECNFWCPCRGSCPMCTHYTQ 167
DB 98 NL-----YMTDPTLMSYVQSWYENEDFVYGVGAK-----PNSAVG---HYQLVW 142
QY 168 ATTKIKGCAVNTCRKKTWGEVWENAVYF--VCNYSKPGNI--GEAPYKNGRSCSECP 224
DB 143 YSSBKICGCIAYCPNOD-----DNLKTFYVCHYCPGNNVWKKSPYQGTGTCASCP-- 193
QY 225 YGSGCRNMLC 234
DB 194 --NNCENGIC 201

RESULT 10
HELO_HELHO
ID HELO_HELHO STANDARD; PRT: 242 AA.
AC Q91055;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Heloderma precursor (HLTX).
OS Heloderma horridum horridum (Mexican beaded lizard).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;
OC Lepidosauria; Squamata; Scleroglossa; Anguimorpha; Helodermatidae;
OC Heloderma.
OX NCBI_TaxID=8552;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RC TISSUE=Salivary gland;
RX MEDLINE=95375162; PubMed=7647234;
RA Morrisette J., Kraetzschmar J., Haendler B., El-Hayek R.,
RA Mochoa-Morales J., Martin B.M., Patel J.R., Moss R.L.,
RT "Primary structure and properties of helodermin, a peptide toxin
RT that blocks ryanodine receptors.";
RL Biophys. J. 68:2280-2288(1995).
CC SEQUENCE OF 20-39, AND CHARACTERIZATION.

```


RC TISSUE-Venom;
 RA PubMed-1693019;
 RX Mochca-Morales J., Martin B.M., Possani L.D.;
 RT "Isolation and characterization of heliothermine, a novel toxin from
 RT Heliothema horridum horridum (Mexican beaded lizard) venom.";
 RL Toxicon 28:299-309(1990).
 CC -1- FUNCTION: Toxin that blocks ryanodine receptors. It is toxic to
 CC mice; it causes lethargy, partial paralysis of rear limbs and
 CC lowering of body temperature.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- TISSUE SPECIFICITY: Produced by the venomous gland
 CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MAMMALIAN SCP/TPX1;
 CC INSECTS AG3/AG5; FUNGI SC7/SC14 AND PLANTS PR-1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U13619; AAC59730.1; -
 CC HSSP: P04284; ICEE.
 CC Interpro: IPR001283; SCP.
 CC Pfam: PF00188; SCP. 1.
 CC PRINTS: PR00837; V5TPXLIKE.
 CC SMART: SM00198; SCP. 1.
 CC PROSITE: PS01009; SCP_AG5_PRL_SC7_1; 1.
 CC PROSITE: PS01010; SCP_AG5_PRL_SC7_2; FALSE_NEG.
 CC Calcium channel inhibitor; Neurotoxin; Toxin; Signal.
 CC SIGNAL 1 19
 CC CHAIN 20 242
 CC SEQUENCE 242 AA; 27493 MW; 0E183FC2F925DF3C CRC64;

 Query Match 9.3%; Score 257.5; DB 1; Length 242;
 Best Local Similarity 31.5%; Pred. No. 1.8e-13;
 Matches 64; Conservative 28; Mismatches 80; Indels 31; Gaps 6;
 QY 54 EDEELIMLNKRGQYQPOASNMETWDELEKSAAMASOCIMRHGP---TSLIYS 109
 DB 34 DOGETEDKNNLRIVEPASNMLKMTWSNKIAQNSANCTLEHTEKEEPTIDGVE 93
 QY 110 IGNLGAHMGYRSPGFHVSQYDEVYDVTYPYPSSECPNCPBERCSGP-----MCTHYTQ 164
 DB 94 CGENLFPSSAPY-TWSAIAIONWDEKRYERFN-----GPTAQNWNIHGYTQ 139
 QY 165 IYVATTNKGICAVNTCKRMVGEWENAVYFCNYSKPGNWIGE--APYKNGRPSCECP 222
 DB 140 VVWYRSYELGCAIAYCPDQPTY-----KYQYQYCPGGINRSKRYTPYSIGPGCDP 193
 QY 223 PSYGGSCNNLCYRETYTPKPE 245
 DB 194 DACDNGICTNPKQNDYNNCPD 216

 RESULT 11
 VAS_VESVU STANDARD; PRT; 227 AA.
 AC 005110; 090B91;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Venom allergen 5 precursor (Antigen 5) (AG5) (Allergen Ves v 5) (Ves v
 V).
 OS Vespaula vulgaris (Yellow jacket) (Wasp).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata;
 OC Vespidae; Vespinae; Vespinae; Vespa.
 OC NCBI_TaxId=7454;
 RN [1]
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.

RC TISSUE-Venom;
 RA MEDLINE-93203603; PubMed-8454859;
 RX Lu G., Villalba M., Coscia M.R., Hoffman D.R., King T.P.;
 RT "Sequence analysis and antigenic cross-reactivity of a venom
 RT allergen, antigen 5, from hornets, wasps, and yellow jackets.";
 RL J. Immunol. 150:2823-2830(1993).
 CC [2]
 CC SEQUENCE OF 24-227 FROM N.A.
 CC "Molecular cloning of a genomic sequence from the venom allergen
 CC antigen 5 from Vespa vulgaris.";
 CC Submitted (MAR-1999) to the EMBL/GenBank/DBI databases.
 CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MAMMALIAN SCP/TPX1;
 CC INSECTS AG3/AG5; FUNGI SC7/SC14 AND PLANTS PR-1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M98858; AAA30333.1; -
 CC HSSP: A0238849; CAB42887.1; -
 CC Interpro: IPR001283; SCP.
 CC Pfam: PF00188; SCP. 1.
 CC PRINTS: PR00837; V5TPXLIKE.
 CC SMART: SM00198; SCP. 1.
 CC PROSITE: PS01009; SCP_AG5_PRL_SC7_1; 1.
 CC PROSITE: PS01010; SCP_AG5_PRL_SC7_2; 1.
 CC Venom; Allergen; Signal.
 CC SIGNAL 1 23
 CC CHAIN 24 227
 CC SEQUENCE 227 AA; 25798 MW; 99E9813740A66F55 CRC64;

 Query Match 9.3%; Score 256; DB 1; Length 227;
 Best Local Similarity 30.5%; Pred. No. 2.2e-13;
 Matches 71; Conservative 33; Mismatches 73; Indels 56; Gaps 12;
 QY 1 MSCVIGVITPLGLFLVCGSOGYLLPNVTLLLELLSKYQHNSHSRVRAIPREDKEIL 60
 DB 29 IKCLKGV-----HTAC-KYGSGLKPNCG--NKVVSV-----GLTKQKQDIL 68
 QY 61 MLHNKLRGV-----QPOASNMETWDELEKSAAMASOCIMRHGP---TSLIYS 108
 DB 69 KEHNDFRQKILARGLTRGNPNPGPPAKNMKNLVMDELAYVAQWVANOCQGHDCRDVA 128
 QY 109 --SIGONL--GAHWGRFRSPGFHVSQYDEVYDVTYPYPSSECPNCPBERCSGP---MCT 160
 DB 129 KYQYQYVALTGSTRAKDDEPKYKLMWDEYKDY-----NP--KKKTSNGDPLTKTG 178
 QY 161 IYVATTNKGICAVNTCKRMVGEWENAVYFCNYSKPGNWIGEAPYK 213
 DB 179 HYTGWVWNTKEVGGC-----SIKYIQKMKH-KLYVNCYSGSNFMNEELIYQ 225

 RESULT 12
 VAS_VESFL STANDARD; PRT; 204 AA.
 AC P35783;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)

```

RC TISSUE=Venom;
RX MEDLINE=94044316; PubMed=8227862;
RA Hoffman D.R.;
RT "Allergens in Hymenoptera venom. XXV: The amino acid sequences of
RT antigen 5 molecules and the structural basis of antigenic cross-
RT reactivity.";
RL J. Allergy Clin. Immunol. 92:707-716(1993).
CC -1 SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MAMMALIAN SCF/TEXT1;
CC INSECTS AG3/AG5; FUNGI SC7/SC14 AND PLANTS PR-1.
DR PIR: C44522; C44522.
DR PIR: C44583; C44583.
DR HSP: P04284; ICFE.
DR InterPro: IPR001283; SCP.
DR Pfam: PF00188; SCP. 1.
DR PRINTS: PRO0837; V5PEXLIKE.
DR SMART: SM00198; SCP. 1.
DR PROSITE: PS01009; SCP_AG5_PRL_SC7_1; 1.
DR PROSITE: PS01010; SCP_AG5_PRL_SC7_2; 1.
DR Venom; Allergen.
FT DISULFD 4 17 BY SIMILARITY.
FT DISULFD 8 101 BY SIMILARITY.
FT DISULFD 26 94 BY SIMILARITY.
FT DISULFD 170 187 BY SIMILARITY.
SQ SEQUENCE 204 AA; 23317 MW; 85BD971066C7D7C8 CRC64;

Query Match 8.8%; Score 244; DB 1; Length 204;
Best Local Similarity 30.4%; Pred. No. 1.7e-12;
Matches 73; Conservative 29; Mismatches 66; Indels 70; Gaps 13.

QY 1 MSCVGVYIPGLFLFLVCGSGCYLLPNVLTLELLSKYQHNSHSRVRRAIPREDKEITL 60
Db 6 IKCLGIGV-----HTAC-KYGSLEKPCNG--NKIVSY-::: -:::1:::1:
QY 61 MHNKLRGV-----QPASNNEYVTMDDELEKSAAMASOCINHGFTSLIV 108
Db 46 KEHNDPFQRIARGLETGRCNPGCPQAPAKNNKNIWNDELAAYVAQVANAOCQYGHDCRIVA 105
QY 109 S--IGQNT--GAHGRYRSPGFHVQSWYDEKYDTYYP-----SECPWPCPERCSGPMG 159
Db 106 KYPVQGNALVGTSTADKYNDPVLKVMMEDEYKDYIN--PKKKEENN-----FKKI 154
QY 160 THYQIWAATNKKIGCAVNTCKRKTVMGEV-----WENAVYFVQVNSYSPKGNWIGADPYK 213
Db 155 GHYTDVWANTKEIGC-----GSIKYTONEMWKKH-YLVCNYPGSGNFGNEELYQ 202

RESULT 14
VA3_SOLIN STANDARD; PRT; 234 AA.
AC P35778; O16135; Q9TWZ2.
DT 01-JUN-1994 (Rel. 29, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Venom allergen III precursor (Allergen Sol 1 3) (Sol 1 III).
OS Solenopsis invicta (Red imported fire ant).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata;
OC Formicidae; Myrmicinae; Solenopsis.
OX NCBI_TaxID=13686;
RN 11
RP SEQUENCE FROM N.A.
RC TISSUE=Venom.
RA Hoffman D.R., Farrar D., Schmidt M., McConnell T.J.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE OF 23-234 FROM N.A.
RX MEDLINE=96051059; PubMed=8588684;
RA Hoffman D.R.;
RT "Fire ant venom allergy."
RT Allergy 50:535-544(1995).
RN 13
RP SEQUENCE OF 23-234.

```

```

RX TISSUE-Venom;
RA MEDLINE-93139387; PubMed-8423273;
RA Hoffman D.R.;
RT "Allergens in Hymenoptera venom XXIV: the amino acid sequences of
RT imported fire ant venom allergens Sol I II, Sol I III, and Sol I
RT IV";
RL J. Allergy Clin. Immunol. 91:71-78(1993).
RN [4]
RP PARTIAL SEQUENCE OF 23-45.
RC TISSUE-Venom;
RA MEDLINE-90285439; PubMed-2355158;
RA Hoffman D.R., Smith A.M., Schmidt M., Moffitt J.E., Guralnick M.;
RT "Allergens in Hymenoptera venom. XXII. Comparison of venoms from two
RT species of imported fire ants, Solenopsis invicta and richteri.";
RL J. Allergy Clin. Immunol. 85:988-996(1990).
CC -1- DISEASE: THE MOST COMMON CAUSE OF INSECT VENOM ALLERGY IN THE
CC SOUTH-EASTERN UNITED STATES IS THE IMPORTED FIRE ANT.
CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MAMMALIAN SCF/TPX1;
CC INSECTS AG3/AG5; FUNGI SC7/SC14 AND PLANTS PR-1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AF012919; AAB65434.1; -
DR PIR; B37330; B37330.
DR PIR; C44582; C44582.
DR HSSP; P04284; 1CFE.
DR InterPro; IP001283; SCP.
DR Pfam; PF00188; SCP; 1.
DR PRINTS; PR00837; V5TPXLIKE.
DR SMART; SM00198; SCP; 1.
DR PROSITE; PS01009; SCP_AGS_PRL_SCF_1; 1.
DR PROSITE; PS01010; SCP_AGS_PRL_SCF_2; 1.
KW Venom; Allergen; Signal.
FT SIGNAL 1 22
FT CHAIN 23 234
FT DISULFID 26 41
FT DISULFID 31 125
FT DISULFID 52 118
FT DISULFID 198 216
FT CONFLICT 183 216
FT CONFLICT 194 194
FT CONFLICT 199 199
SQ SEQUENCE 234 AA; 26351 MW; 539F510B59941D83 CRC64;

Query Match 8.8%; Score 243.5; DB 1; Length 234;
Best Local Similarity 35.0%; Pred. No. 2.2e-12;
Matches 63; Conservative 23; Mismatches 61; Indels 33; Gaps 7;
OY 55 DEKELIMLNKRGV-----OPASNMEYMTWDELLEKSAAMASQCTWEGH 102
DB 64 EKDALVKNHNEIRVAVSGKEKRGNGPQPPAVKMPNLTDELATIAORMANOCCTFEHD 123
OY 103 PLSL--VSIGONLGA--HMGRRS--PGFHVQSWDEVKDY---TYPPSECNPCPER 153
DB 124 ACRNVERAFCGNINATSSCKNKNSTPMILLWYNEVDFDNRISSPDSNIIIM--- 180
OY 154 CSGPACTHYTQIWMATNKGCAVNTCRKMTWGEWENAVYFVCNYSKPMIGEPAPYK 213
DB 181 ----KVGHTYQIWMATNKGCAVNTCRKMTWGEWENAVYFVCNYSKPMIGEPAPYK 231

```

RESULT 15
 VAS_VESMC STANDARD: PRT: 204 AA.
 ID VAS_VESMC
 AC P35760;
 DT 01-JUN-1994 (rel. 29, Created)

```

DT 01-JUN-1994 (rel. 29, last sequence update)
DT 01-NOV-1997 (rel. 35, last annotation update)
DE Venom allergen 5 (Antigen 5) (AG5) (Allergen Ves m 5) (Ves m V).
OS Vespa maculifrons (Eastern yellow jacket) (Wasp).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata;
OC Vespoidea; Vespidae; Vespinae; Vespa.
OX NCBI_TaxId=7453;
RN [1]
RP SEQUENCE.
RC TISSUE-Venom;
RA MEDLINE-93203603; PubMed-8454859;
RA Lu G., Villalba M., Coscia M.R., Hoffman D.R., King T.P.;
RT "Sequence analysis and antigenic cross-reactivity of a venom
RT allergen, antigen 5, from hornets, wasps, and yellow jackets.";
RL J. Immunol. 150:2823-2830(1993).
CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MAMMALIAN SCF/TPX1;
CC INSECTS AG3/AG5; FUNGI SC7/SC14 AND PLANTS PR-1.
CC -----
DR PIR; B37329; B37329.
DR HSSP; P04284; 1CFE.
DR InterPro; IP001283; SCP.
DR Pfam; PF00188; SCP; 1.
DR PRINTS; PR00837; V5TPXLIKE.
DR SMART; SM00198; SCP; 1.
DR PROSITE; PS01009; SCP_AGS_PRL_SCF_1; 1.
DR PROSITE; PS01010; SCP_AGS_PRL_SCF_2; 1.
KW Venom; Allergen.
FT DISULFID 4 17
FT DISULFID 8 101
FT DISULFID 26 94
FT DISULFID 170 187
FT DISULFID 170 187
SQ SEQUENCE 204 AA; 23332 MW; 8B68A94C41390311 CRC64;

Query Match 8.7%; Score 240; DB 1; Length 204;
Best Local Similarity 29.6%; Pred. No. 3.6e-12;
Matches 68; Conservative 27; Mismatches 85; Indels 50; Gaps 9;
OY 1 MSCGLGVIPILGLFVCGSQGYLLPNVTLLEELLSKYQHNEHSRYRAIPREDKEIL 60
DB 6 IKLKAGV-----HNAC-KYSLKPNK-----GNKKYVSYLTKQEKODIL 45
OY 61 MLNKNKRGV-----OPASNMEYMTWDELLEKSAAMASQCTWEGHPSLLV 108
DB 46 KEHNDPROKILARGLERGNPGPQPPAKNMKNLWSEDELAIVIAQWMANOCQYGHDCRVA 105
OY 109 --SIGONL--GAMHGRYSRPGFHVQSWDEVKDYTPYPSSECNPCPERSGPACTHYT 163
DB 106 KYVGONVALTGSTAATYNDPVKLVKMWDEVDKDYDPKPKKFSNNFL-----KIGHY 158
OY 164 QIWMATNKGCAVNTCRKMTWGEWENAVYFVCNYSKPMIGEPAPYK 213
DB 159 QMWMATNKGCG-----SIKYQIWMHKNH-LYVCNYSKPMIGEPAPYK 202

```

Search completed: May 25, 2002, 14:39:57
 Job time: 662 sec

SEA ID NO: 2

AC NO: AAY41738, Database: A-Geneg-032802

File Copy

117 HMGTRAGGFGHVSQMYDEVKDYTPYPSECNPCPCSGPMCTHYTOYATNTKICGA 176
 3 nmgvyrpftivgwydevkdytppsecnmpcpcsgpmcthytqivattkicga 62
 177 VNTCKRKTATGSEWENAVYFCNYSPPKGNWIGEPYKNGRCPSPSYGSGCRNNLCYR 236
 63 vntckrktatgsewenvayfvcnysppkgnwlgearpkngrpsecpspsycsrnlcyr 122
 237 EETYPKRETDMMNENETAPPEENHVMLOPRVMPRTPKRTSAVNYMTQVVRCDTKMKD 296
 123 eetyprkretdmnenetapipreenhvmloprvmprtpkrtksavnymtqvvrctdkmkd 182
 297 RCKGSTCNRYOCFACCLNHNKATIFGSLFTSSSSICRAIHYGIIDDKGGLVDITRNGVY 356
 183 rckgstcnryocfaccclnhnkatifgslfysessicraahygilddkgylditrngv 242
 357 PFTVYSEHGVOSLSKTPRSSSEMYSKYVODLDCYTTVAOLCPFEKPTATHCPRIHCPAH 416
 243 pftvserhgvoslsktpsssemyskvovodldcyltvaqlcpfekpattchprihcpah 302
 417 CDEPSTAPVFTNTYATPSTGICKTAVHAGVLSNMSGDVPDVPVKKKTYVGSIRNGY 476
 303 cdepstapvftntyatpstgicktavhagvlsnmsgdvpdvpdkkktyvgsirngy 362
 477 QSESLGTPRDGKAFRIFAVRQ 497
 363 qseslgtprdgkafirifavrq 383

RESULT 7
 AAB93979 standard; Protein; 371 AA.
 AAB93979:
 26-JUN-2001 (first entry)
 Human protein sequence SEQ ID NO:14059.
 Human; primer; detection; diagnosis; antitense therapy; gene therapy.
 Homo sapiens.
 EP1074617-A2.
 07-FEB-2001.
 28-JUL-2000; 2000EP-0116126.
 29-JUL-1999; 99JP-0248036.
 27-AUG-1999; 99JP-0300253.
 11-JAN-2000; 2000JP-0118776.
 02-MAY-2000; 2000JP-0183767.
 09-JUN-2000; 2000JP-0241899.
 (HELI-) HELIX RES INST.
 Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 Ishii S, Sugiyama T, Wakematsu A, Nagai K, Otsuki T;
 WPI; 2001-318749/34.
 Primer sets for synthesizing polynucleotides, particularly the 5602
 full-length cDNAs defined in the specification, and for the detection
 and/or diagnosis of the abnormality of the proteins encoded by the
 full-length cDNAs -
 Claim 8; SEQ ID 14059; 2537bp + CD ROM; English.
 The present invention describes primer sets for synthesizing 5602
 full-length cDNAs defined in the specification. Where a primer set
 comprises: (a) an oligo-dT primer and an oligonucleotide complementary

to the complementary strand of a polynucleotide which comprises one of
 the 5602 nucleotide sequences defined in the specification, where the
 oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 of an oligonucleotide comprising a sequence complementary to the
 complementary strand of a polynucleotide which comprises a 5'-end
 sequence and an oligonucleotide comprising a sequence complementary to a
 polynucleotide which comprises a 3'-end sequence where the
 oligonucleotide comprises at least 15 nucleotides and the combination of
 the 5'-end sequence/3'-end sequence is selected from those defined in
 the specification. The primer sets can be used in antisense therapy and
 in gene therapy. The primers are useful for synthesizing polynucleotides,
 particularly full-length cDNAs. The primers are also useful for the
 detection and/or diagnosis of the abnormality of the proteins encoded by
 the full-length cDNAs. The primers allow obtaining of the full-length
 cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 represent oligonucleotides, all of which are used in the exemplification
 of the present invention.

Sequence 371 AA:

Query Match 75.1%; Score (2071); DB 22; Length 371;
 Best Local Similarity 99.7%; Pred. No. 2.7e-169;
 Matches 369; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

1 MSCVIGVPIPLGLFLVCGSGYLIPVTLLEELSKYOHNESHRYRRAIPREDKEIL 60
 1 mscvlgvpiplglflvcsqgyllpvtlleelkskyqhneshrrraipredkeil 60
 61 MLHNKLGQVQPOASNNMYTMDLEKSAANAASQCIWEHGPSLSLVSGONTGAMGR 120
 61 mlhnklrgvqpqasnmeytmdleksaanaasqciwehgpsllsvsgnlgamgr 120
 121 YRSPGFHVOSMYDEVKDYTPYPSECNPCPCSGPMCTHYTOYATNTKICAVNTC 180
 121 yrspgfhwosmydevkdytppsecnmpcpcsgpmcthytqivattkicavntc 180
 121 yrspgfhwosmydevkdytppsecnmpcpcsgpmcthytqivattkicavntc 180
 181 RKMTVGEWENAVYFCNYSPPKGNWIGEPYKNGRCPSPSYGSGCRNNLCYRETY 240
 181 rkmtvgewenvayfvcnysppkgnwlgearpkngrpsecpspsycsrnlcyreety 240
 181 rkmtvgewenvayfvcnysppkgnwlgearpkngrpsecpspsycsrnlcyreety 240
 241 TPKEPTDMENETAPPEENHVMLOPRVMPRTPKRTSAVNYMTQVVRCDTKMKDRCG 300
 241 tpkpetdmnenetapipreenhvmloprvmprtpkrtksavnymtqvvrctdkmkdrckg 300
 301 STCNRYOCFACCLNHNKATIFGSLFTSSSSICRAIHYGIIDDKGGLVDITRNGVY 360
 301 stcnryocfaccclnhnkatifgslfysessicraahygilddkgylditrngvypfv 360
 361 KSERHGVOSLSKTPRSSSEMYSKYVODLDCYTTVAOLCPFEKPTATHCPRIHCPAH 416
 361 kserhgvoslsktpsssemyskvovodldcyltvaqlcpfekpattchprihcpah 416
 361 kserhgvoslsktpsssemyskvovodldcyltvaqlcpfekpattchprihcpah 416

RESULT 8

AAY41738 standard; Protein; 500 AA.

AAY41738:

07-DEC-1999 (first entry)

Human PRO541 protein sequence.

Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;
 probe; blood coagulation disorder; cancer; cellular adhesion disorder;
 secreted protein; transmembrane protein.

Homo sapiens.

MO9946281-A2.





OY	421	ccctcccccagacgtgttcgaacaccccttggtgtccagagagaggtgcctcggtcgtatgtgcag	480
Db	626	ccctcccccagacgtgttcgaacaccccttggtgtccagagagaggtgcctcggtcgtatgtgcag	685
OY	481	cactacacacagataagtttgggccacacacacacagatcgtgtgtctgttgcagacacctgc	540
Db	686	cactacacacagataagtttgggccacacacacacagatcgtgtgtctgttgcagacacctgc	745
OY	541	cggaaagatgactgtctggggagaaagtttggagaaagcggctacttctgtcgaattat	600
Db	746	cggaaagatgactgtctggggagaaagtttggagaaagcggctacttctgtcgaattat	805
OY	601	tctccaaagvggaaactggtattgtgaagaagccctctacaaagatatggccggccctctag	660
Db	806	tctccaaagvggaaactggtattgtgaagaagccctctacaaagatatggccggccctctag	865
OY	661	tgcccaacccagctcttgagagcagctgcaggaacaaactgtgtataccggagaaagaaactac	720
Db	866	tgcccaacccagctcttgagagcagctgcaggaacaaactgtgtataccggagaaagaaactac	925
OY	721	actccaaaaccttgaaagcgcgacgataabtgaaagtggaaacggcttccattcctgaagaa	780
Db	926	actccaaaaccttgaaagcgcgacgataabtgaaagtggaaacggcttccattcctgaagaa	985
OY	781	aaccatgtttggtccccaacccgaggtgtatgtaaacccaacgaagcccaagaacacctctcg	840
Db	986	aaccatgtttggtccccaacccgaggtgtatgtaaacccaacgaagcccaagaacacctctcg	1045
OY	841	gtcaactctatagacccaagctcgtcagatgtgtacacacaaagtgaagagacagtgtcaaaag	900
Db	1046	gtcaactctatagacccaagctcgtcagatgtgtacacacaaagtgaagagacagtgtcaaaag	1105
OY	901	tccaacgtgtaaacaggtaccaggtgtcccaacagcgtgtctgaacccaagcggaaagcttt	960
Db	1106	tccaacgtgtaaacaggtaccaggtgtcccaacagcgtgtctgaacccaagcggaaagcttt	1165
OY	961	ggaagctgtgtctatagaagctcgtctagcatatgcgcgcgcgcacatccactacagagatc	1020
Db	1166	ggaagctgtgtctatagaagctcgtctagcatatgcgcgcgcgcacatccactacagagatc	1225
OY	1021	ctggatatcaagvggagccgcgtgtgatatatacagaagaaggaaagttccctctctctgt	1080
Db	1226	ctggatatcaagvggagccgcgtgtgatatatacagaagaaggaaagttccctctctctgt	1285
OY	1081	aagctctgagagacacggcgtgtgcagctccctcagaagaatacaaaccttcacgcactatcat	1140
Db	1286	aagctctgagagacacggcgtgtgcagctccctcagaagaatacaaaccttcacgcactatcat	1345
OY	1141	gtgtccaaaagtgtgaagtgacgggatttggactgttaacgaacgcgtgtgtcagctgtgcgg	1200
Db	1346	gtgtccaaaagtgtgaagtgacgggatttggactgttaacgaacgcgtgtgtcagctgtgcgg	1405
OY	1201	tttgtaaaagccagaaactcactctgcacaaagatatctgtgtccggacacacgtgcagaagaa	1260
Db	1406	tttgtaaaagccagaaactcactctgcacaaagatatctgtgtccggacacacgtgcagaagaa	1465
OY	1261	ctctctcactcgggtcctccggtgttttgtaaacacacatctatgacagataactcaagcacttgc	1320
Db	1466	ctctctcactcgggtcctccggtgttttgtaaacacacatctatgacagataactcaagcacttgc	1525
OY	1321	aagacagactgtgcacggggagatcatcagaacaaagaaagtgggggtgaaagtgtgaagctgtg	1380
Db	1526	aagacagactgtgcacggggagatcatcagaacaaagaaagtgggggtgaaagtgtgaagctgtg	1585
OY	1381	cccgctgtaataaaaagaaagacctacgtgtggctcgtctcagaagaaatgaaagcttcaagctgaagc	1440
Db	1586	cccgctgtaataaaaagaaagacctacgtgtggctcgtctcagaagaaatgaaagcttcaagctgaagc	1645
OY	1441	ctggggaactcctcgggattggaagaagccttcggagatttgcgtcgaagcag	1491
Db	1646	ctggggaactcctcgggattggaagaagccttcggagatttgcgtcgaagcag	1696

RESULT	7	
ID	AAD17766	standard; DNA; 2403 BP.
XX		
XX	AAD17766;	
XX		
DT	10-DEC-2001	(first entry)
XX		
XX		
XX	Human novel trypsin inhibitor-like protein, NOV-4d encoding DNA.	
KW	Human; NOV-X protein; KIA1233-like protein; STE20-like protein; tumour	
KW	trypsin inhibitor-like protein; gene therapy; haematopoietic illness;	
KW	immunological disorder; neurodegenerative disorder; Alzheimer's disease	
KW	Parkinson's disease; immunomodulatory; pharmacogenomic; haemostatic;	
KW	human immunodeficiency virus; HIV; fertility disorder; neuroprotective;	
KW	cytostatic; neutrotropic; anti-fertility; cancer; NOV-4d protein; ds.	
XX		
OS	Homo sapiens.	
XX		
FT	Key	Location/Qualifiers
FT	CDS	206..1702
FT		/*tag= a
FT		/product= "Human novel trypsin inhibitor-like protein,
FT		NOV-4d"
XX		
PN	W0200162928-A2.	
XX		
XX	30-AUG-2001.	
PD		
XX		
PF	26-FEB-2001; 2001WO-US06151.	
XX		
PR	25-FEB-2000; 2000US-0184951.	
PR	28-FEB-2000; 2000US-0185548.	
PR	01-MAR-2000; 2000US-0185967.	
PR	18-APR-2000; 2000US-0197723.	
PR	27-APR-2000; 2000US-0199957.	
PR	23-FEB-2001; 2001US-0789390.	
XX		
PA	(CUBRA-) CUBRA GEN CORP.	
PI	Vernet CAM, Fernandes E, Shinkets RA, MacDougall J, Spaderna SK;	
DR	WPI; 2001-582051/65.	
DR	P-PSDB; AAE17766.	
XX		
PT	New isolated KIA1233-like, STE20-like, or trypsin inhibitor-like	
PT	polypeptide for diagnosing and treating pathological disorders, such as	
PT	Parkinson's disease and for use in pharmacogenomics -	
XX		
PS	Claim 9; Page 87-88; 189pp; English.	
XX		
CC	The invention relates to novel human polypeptides referred as NOV-X	
CC	and their corresponding nucleic acid sequences. NOV-X collectively	
CC	include NOV-1, NOV-2a and NOV-2b which are novel KIA1233-like	
CC	polypeptides, NOV-3a, NOV-3b, NOV-3c and NOV-3d which are novel	
CC	STE20-like polypeptides and NOV-4a, NOV-4b, NOV-4c, NOV-4d and NOV-4e	
CC	which are novel trypsin inhibitor-like polypeptides. NOV-X is used to	
CC	identify a potential therapeutic agent that can modulate its activity	
CC	and can be used for treating a pathology related to aberrant expression	
CC	or aberrant physiological interactions of NOV-X. NOV-X or its DNA is	
CC	used to determine the presence or predisposition to a disease associated	
CC	with altered levels of NOV-X. NOV-X, its DNA and its antibody are used	
CC	to treat or prevent a pathology associated with NOV-X. The pathological	
CC	states that can be treated or prevented are haematopoietic, cancer,	
CC	immunological, tumour, neurodegenerative (e.g. Alzheimer's and	
CC	Parkinson's disease), human immunodeficiency virus (HIV) illness and	
CC	fertility disorders. NOV-X and its DNA are used in pharmacogenomics for	
CC	predictive medicine. NOV-X DNA is used in gene therapy. The present	
CC	sequence is a DNA encoding human novel trypsin inhibitor-like protein,	
CC	NOV-4d.	
XX		
Sequence	2403 BP; 558 A; 646 C; 685 G; 514 T; 0 other;	

Tue May 28 16:10:05 2002

us-09-667-380a-2.rspt

Page 1

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 25, 2002, 14:27:40 ; Search time 122.23 Seconds

(without alignments)
703.416 Million cell updates/sec

Title: US-09-667-380a-2

Sequence: 1 MSCVIGVTPGLFLVCGS.....SESLGTPRDGKAFRIFAVNG 497

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database :

- 1: SPREMBL_19:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_ricent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriopl:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2756	99.9	497	4 Q9H0B8	Q9H0B8 homo sapien
2	2071	75.1	371	4 Q96K61	Q96K61 homo sapien
3	1987.5	72.0	434	11 Q9D2R3	Q9D2R3 mus musculu
4	1640.5	59.5	500	4 Q9H336	Q9H336 homo sapien
5	1613	58.5	523	13 Q985T5	Q985T5 gallus galli
6	1536.5	55.7	507	11 Q99M06	Q99M06 mus musculu
7	968	35.1	188	11 Q920U6	Q920U6 rattus norv
8	758.5	27.5	258	13 Q985T6	Q985T6 gallus galli
9	740	26.8	258	11 Q99M07	Q99M07 mus musculu
10	739.5	26.8	258	4 Q43692	Q43692 homo sapien
11	646.5	23.4	253	4 Q9H3Y0	Q9H3Y0 homo sapien
12	639	23.2	120	4 Q96IR1	Q96IR1 homo sapien
13	357	12.9	415	5 Q44228	Q44228 halocynthia
14	355	12.9	266	4 Q969K2	Q969K2 homo sapien
15	349	12.6	489	11 Q9J056	Q9J056 mus musculu
16	349	12.6	489	11 Q9ET66	Q9ET66 mus musculu

17	323.5	11.7	236	11 Q9DAG6	Q9DAG6 mus musculu
18	319.5	11.6	255	11 Q9CWD1	Q9CWD1 mus musculu
19	301	10.9	244	11 Q91XA3	Q91XA3 mus musculu
20	294	10.7	332	11 Q9CQ35	Q9CQ35 mus musculu
21	293.5	10.6	203	4 Q9H108	Q9H108 homo sapien
22	292.5	10.6	203	4 Q9H106	Q9H106 homo sapien
23	286	10.4	380	5 Q9EYF2	Q9EYF2 drosophila
24	286	10.4	392	5 Q960R5	Q960R5 drosophila
25	273.5	9.9	301	5 Q950F6	Q950F6 caenorhabdi
26	273	9.9	220	5 Q16854	Q16854 onchocerca
27	273	9.9	243	11 Q9R1L4	Q9R1L4 rattus norv
28	268.5	9.7	243	11 Q88205	Q88205 rattus norv
29	254	9.2	334	6 Q9BE36	Q9BE36 macaca fasc
30	240	8.7	165	6 Q77720	Q77720 equus caball
31	239	8.7	220	5 Q97149	Q97149 wuchereria
32	239	8.7	250	11 Q9D259	Q9D259 mus musculu
33	237	8.6	196	5 Q62507	Q62507 caenorhabdi
34	235	8.5	220	5 Q4932	Q4932 brugia mala
35	234.5	8.5	424	5 Q9X241	Q9X241 necator ame
36	234.5	8.5	424	5 Q9X241	Q9X241 onchocerca
37	234	8.5	220	5 Q4931	Q4931 onchocerca
38	234	8.5	249	6 Q9XSD3	Q9XSD3 macaca mula
39	233	8.4	176	10 Q93186	Q93186 arabidopsis
40	233	8.4	212	5 Q9B1D5	Q9B1D5 heterodera
41	232	8.4	212	5 Q96764	Q96764 caenorhabdi
42	231.5	8.4	208	5 Q18543	Q18543 caenorhabdi
43	231.5	8.4	217	6 Q77719	Q77719 equus caball
44	230	8.3	190	10 Q95W05	Q95W05 arabidopsis
45	230	8.3	212	5 Q20609	Q20609 caenorhabdi

ALIGNMENTS

RESULT	1	PRELIMINARY	PRT	497 AA.
Q9H0B8				
ID	Q9H0B8			
AC	Q9H0B8			
DT	01-MAR-2001 (TREMREL. 16, Created)			
DT	01-MAR-2001 (TREMREL. 16, Last sequence update)			
DE	01-DEC-2001 (TREMREL. 19, Last annotation update)			
DE	HYPOTHEICAL 55.9 KDA PROTEIN.			
GN	DREFZ43AB044.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=TESTIS;			
RA	Wambutt R., Heubner D., Mewes H.W., Gassenhuber J., Wiemann S.;			
RL	Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AL136861; CAB66795.1; ..			
DR	HSP; P04284; ICFE.			
DR	InterPro: IPR004043; ICDL.			
DR	InterPro: IPR001283; SCP.			
DR	Pfam: PF00188; SCP; 1.			
DR	PRINTS: PR00837; VSTFXLKE.			
DR	SMART; SM00198; SCP; 1.			
KW	PROSITE; PS01010; SCP_AGS_PRL_SCT_2; 1.			
DR	Hypothetical protein.			
SQ	SEQUENCE 497 AA; 55919 MW; 1E7ADDCAB37CA4ED CRC64;			

Query Match 99.9%; Score 2756; DB 4; Length 497;
Best Local Similarity 99.8%; Pred. No. 3.5e-259;
Matches 496; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MSCVIGVTPGLFLVCGSOGYLLPNTLLLEILLKYOHNSHRSVRRAIPREDKEITL 60
Db	1	MSCVIGVTPGLFLVCGSOGYLLPNTLLLEILLKYOHNSHRSVRRAIPREDKEITL 60
Qy	61	MLHNLKRGVOVPOASNMETMTWDELKSAANASQCIWEHGPTSLVSGICNGAHMGR 120

Db 61 MLNKLKGOVPOASNNEMTMDLEKSAAMASQCIWEHPTSLVSGONLGAHGR 120
121 YRSPGFHVOSMYDEVKDYTPYSECPNCPBCSCGPMCHTYQIWAATNKGCAVNTC 180
122 YRSPGFHVOSMYDEVKDYTPYSECPNCPBCSCGPMCHTYQIWAATNKGCAVNTC 180
181 RKMTWGEWENAVYFVCNTPSPKGNWIGEARPYKNGRSCSEPPSYGSGCNRNLCYRETY 240
182 RKMTWGEWENAVYFVCNTPSPKGNWIGEARPYKNGRSCSEPPSYGSGCNRNLCYRETY 240
241 TPKPEDEENNEVEFAPRPEENHNLQPRVMPKPKTSVAVNTQVYVRCDTKMKDRCKG 300
242 TPKPEDEENNEVEFAPRPEENHNLQPRVMPKPKTSVAVNTQVYVRCDTKMKDRCKG 300
301 STCNRYOCPAGCLNHRKAKIFGSLFYESSSSICRAAIHYGLIDKGLVDITRNGKVPFV 360
302 STCNRYOCPAGCLNHRKAKIFGSLFYESSSSICRAAIHYGLIDKGLVDITRNGKVPFV 360
361 KSRHGVOSLSKRPSSSPWVKVQDLDCTYVAQLCPFEKPAHCRIRHCPACRCKDE 420
362 KSRHGVOSLSKRPSSSPWVKVQDLDCTYVAQLCPFEKPAHCRIRHCPACRCKDE 420
421 PSYMAVFEFTNIADTSSICKTAVHAGVISNESGDDVMPVDDKRTYVGSILNAGVOS 480
422 PSYMAVFEFTNIADTSSICKTAVHAGVISNESGDDVMPVDDKRTYVGSILNAGVOS 480
481 LGTPRDGKAFRIFAVRQ 497
482 LGTPRDGKAFRIFAVRQ 497

RESULT 2

AC 096K61 PRELIMINARY: PRT: 371 AA.
ID 096K61
AC 096K61
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CDNA FJ114488.FIS, CLONE YAMAMA100281, WEAKLY SIMILAR TO GLIOMA
DE PAPHOENETIS-RELATED PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MAMMARY GLAND;
RA Itoigai T., Oca F., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuho Y., Sasaki N.,
RA "NGDO human CDNA sequencing project."
RL Submitted (May-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK027395; BAB55081.1;
SQ SEQUENCE 371 AA; 42807 MW; 8D23FBE14F53E85C CRC64;

Query Match 75.1% Score 2071; DB 4; Length 371;
Best Local Similarity 99.7%; Pred. No. 9, 1e-193;
Matches 369; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSCVYGVIPGLLEFVCGSGYLLANTLLEELSTKQHNESRVARAPREDKEITL 60
Db 1 MSCVYGVIPGLLEFVCGSGYLLANTLLEELSTKQHNESRVARAPREDKEITL 60
QY 61 MLNKLKGOVPOASNNEMTMDLEKSAAMASQCIWEHPTSLVSGONLGAHGR 120
Db 61 MLNKLKGOVPOASNNEMTMDLEKSAAMASQCIWEHPTSLVSGONLGAHGR 120
QY 121 YRSPGFHVOSMYDEVKDYTPYSECPNCPBCSCGPMCHTYQIWAATNKGCAVNTC 180
Db 121 YRSPGFHVOSMYDEVKDYTPYSECPNCPBCSCGPMCHTYQIWAATNKGCAVNTC 180

QY 181 RKMTWGEWENAVYFVCNTPSPKGNWIGEARPYKNGRSCSEPPSYGSGCNRNLCYRETY 240
182 RKMTWGEWENAVYFVCNTPSPKGNWIGEARPYKNGRSCSEPPSYGSGCNRNLCYRETY 240
241 TPKPEDEENNEVEFAPRPEENHNLQPRVMPKPKTSVAVNTQVYVRCDTKMKDRCKG 300
242 TPKPEDEENNEVEFAPRPEENHNLQPRVMPKPKTSVAVNTQVYVRCDTKMKDRCKG 300
301 STCNRYOCPAGCLNHRKAKIFGSLFYESSSSICRAAIHYGLIDKGLVDITRNGKVPFV 360
302 STCNRYOCPAGCLNHRKAKIFGSLFYESSSSICRAAIHYGLIDKGLVDITRNGKVPFV 360
361 KSRHGVOSLSKRPSSSPWVKVQDLDCTYVAQLCPFEKPAHCRIRHCPACRCKDE 420
362 KSRHGVOSLSKRPSSSPWVKVQDLDCTYVAQLCPFEKPAHCRIRHCPACRCKDE 420

RESULT 3

ID 09D2R3 PRELIMINARY: PRT: 434 AA.
AC 09D2R3
AC 09D2R3
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE 1810049K24RIK PROTEIN.
GN 1810049K24RIK
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=PANCREAS;
RX MEDLINE=21083660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aikawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana S.,
RA Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fieischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Mikalido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli R., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Funtun M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Kamita M., Lee N.H.,
RA Gustincich S., Hill D., Holtzman M., Hume D.A., Kamita M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momberts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez C., Sato T., Shibata Y.,
RA Sasaki H., Sato K., Schoenbach C., Seva T., Shitaker C., Storch K.F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kontsuki S.,
RA Hayashizaki Y.,
RT "Functional annotation of a full-length mouse CDNA collection."
RL Nature 409:685-690(2001).
DR EMBL: AK019034; BAB31519.1;
DR HSSP: P04284; ICPE.
DR MGD: MGI:1926142; 1810049K24RIK.
DR InterPro: IPR004043; LCCL.
DR InterPro: IPR001283; SCP.
DR Pfam: PF00188; SCP; 1.
DR PRINTS: PR00837; VSTPXLKE.
DR SMART: SM00198; SCP; 1.
DR PROSITE: PS01010; SCP; AG5_PRL_SCT_2;
SQ SEQUENCE 434 AA; 48593 MW; C25067E914647AC0 CRC64;

Query Match 72.0% Score 1987; DB 11; Length 434;
Best Local Similarity 79.6%; Pred. No. 1, 1e-184;
Matches 348; Conservative 37; Mismatches 49; Indels 3; Gaps 1;

QY 61 MLNKLKGOVPOASNNEMTMDLEKSAAMASQCIWEHPTSLVSGONLGAHGR 120
Db 61 MLNKLKGOVPOASNNEMTMDLEKSAAMASQCIWEHPTSLVSGONLGAHGR 120

May

XX Claim 2: Page 27-28; 29pp; English.
 PS The present invention provides the protein and coding sequences of the
 CC novel human trypsin inhibitor like protein NHP. This shows homology to
 CC mammalian trypsin inhibitors. The sequences are useful in disease
 CC diagnosis and treatment, particularly of diseases associated with signal
 CC transduction. The present sequence is the NHP protein.
 XX
 SQ Sequence 497 AA:

Query Match 100.0%; Score 2759; DB 22; Length 497;
 Best Local Similarity 100.0%; Pred. No. 3.6e-228;
 Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSCVLAGVPLGLFLVCSOGYLLPNVTLLEELSKYOHNSHSHVRAIRPREDEEITL 60
 DB 1 mscvlgvplglflfvcsogyllpnvtlleelskqhshsvrrairpredkeell 60
 QY 61 MHNKLRGVQPOASNMEYMTWDELKSAAMASOCIEHGPTSLVSIIGNLGAHWGR 120
 DB 61 mhnklrgvqpoasnmeymtwdelekasaawaasqciwebhptslvsiignlgahwgr 120
 QY 121 YRSPGFHVSQWYDEVKDYTPRPSECNPCPCRCGPMCTHTQIYMATNTKIGCAVNTC 180
 DB 121 yrspgfhwswydevkdytrpsecpncpcrcgpmcthytqlwatkngicavntc 180
 QY 181 RKMATWGEWENAVYFVCNYSFKGNMIGEPYKNGRPSCEPSPYSGSCNNLCYREETY 240
 DB 181 rkmawgewenavfyvcnysfkgnmigepykngprscsepssygsccnnlcyreety 240
 QY 241 TPKEPTDENNETAPIPENHVMLOPRMPTPKTKTSANVTMNVQVCDTAKRCRG 300
 DB 241 tpkptdenneetapipenhvmloprmptrpktktsanvmtmvqvcvdtakmrcrg 300
 QY 301 STCNRYOCPCAGLNNKAKIFGSLFESSSSSICRAAIHYGILDDKGLVDITRNGKVPFY 360
 DB 301 stcnryocpcaglnnkakifgslfeyssssicraaihygildldkglvdtitngkvpfy 360
 QY 361 KSEBRGVQSLSKYKSSSSMWKRVKQVODICTTVAQLCPFEKPAHCPRIHCAHCKDE 420
 DB 361 ksebrgvqslskykssssmwkrvkqvodicttvaqlcpfekpahcprihcapckde 420
 QY 421 PSYMAPVFCTINTYADTSSICTAVNAHQAIVSNESGGVDVMPDPDKKTYGSLRNGVQSSS 480
 DB 421 psymapvfctintyadtssictavnaahqavsnegsgvdvmpdpdkktyvgslnrgvqss 480
 QY 481 LGTPRDGKAFRIFAVRQ 497
 DB 481 lgtprdgkafirifavrq 497

RESULT 2
 ID AAE10616 standard; Protein: 497 AA.

AAE10616;

10-DEC-2001 (first entry)

Human novel trypsin inhibitor-like protein, NOV-4b.

Human; NOV-X protein; KIA1233-like protein; STE20-like protein; tumour;
 trypsin inhibitor-like protein; gene therapy; haematopoietic; illness;
 immunological disorder; neurodegenerative disorder; Alzheimer's disease;
 Parkinson's disease; immunomodulatory; pharmacogenomic; haemostatic;
 human immunodeficiency virus; HIV; fertility disorder; neuroprotective;
 cytosolic; neotropic; anti-fertility; cancer; NOV-4b protein.

Homo sapiens.

Key Location/Qualifiers

FT Peptide 1..22
 FT /label= signal_peptide
 FT Protein 23..497
 FT /note= "Human mature novel trypsin inhibitor-like
 FT protein, NOV-4b"
 FT Domain 195..206
 FT /label= SCP_domain
 PN WO200162928-A2.
 PD 30-AUG-2001.
 XX 26-FEB-2001; 2001WO-US06151.
 XX 25-FEB-2000; 2000US-0184951.
 XX 28-FEB-2000; 2000US-0185548.
 XX 01-MAR-2000; 2000US-0185967.
 XX 18-APR-2000; 2000US-0197723.
 XX 27-APR-2000; 2000US-0199957.
 XX 23-FEB-2001; 2001US-0789390.
 XX (CDRA-) CUBAGEN CORP.
 PI Vernet CAM, Fernandes E, Shinkets RA, Macdougall J, Spaderna SK;
 XX WPI: 2001-582051/65.
 DR N-PSDB: AAD17764.
 PT New isolated KIA1233-like, STE20-like, or trypsin inhibitor-like
 PT polypeptide for diagnosing and treating pathological disorders, such as
 PT Parkinson's disease and for use in pharmacogenomics -
 XX
 PS Claim 1: Page 83; 189pp; English.
 XX
 CC The invention relates to novel human polypeptides referred as NOV-X
 CC and their corresponding nucleic acid sequences. NOV-X collectively
 CC include NOV-1, NOV-2a and NOV-2b which are novel KIA1233-like
 CC polypeptides, NOV-3a, NOV-3b, NOV-3c and NOV-3d which are novel
 CC STE20-like polypeptides and NOV-4a, NOV-4b, NOV-4c, NOV-4d and NOV-4e
 CC which are novel trypsin inhibitor-like polypeptides. NOV-X is used to
 CC identify a potential therapeutic agent that can modulate its activity
 CC and can be used for treating a pathology related to aberrant expression
 CC or aberrant physiological interactions of NOV-X. NOV-X or its DNA is
 CC used to determine the presence or predisposition to a disease associated
 CC with altered levels of NOV-X. NOV-X, its DNA and its antibody are used
 CC to treat or prevent a pathology associated with NOV-X. The pathological
 CC states that can be treated or prevented are haematopoietic, cancer,
 CC immunological, tumour, neurodegenerative (e.g. Alzheimer's and
 CC Parkinson's disease), human immunodeficiency virus (HIV) illness and
 CC fertility disorders. NOV-X and its DNA are used in pharmacogenomics for
 CC predictive medicine. NOV-X DNA is used in gene therapy. The present
 CC sequence is human novel trypsin inhibitor-like protein, NOV-4b.
 XX
 SQ Sequence 497 AA:

Query Match 99.4%; Score 2743; DB 22; Length 497;
 Best Local Similarity 99.6%; Pred. No. 8.6e-227;
 Matches 495; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSCVLAGVPLGLFLVCSOGYLLPNVTLLEELSKYOHNSHSHVRAIRPREDEEITL 60
 DB 1 mscvlgvplglflfvcsogyllpnvtlleelskqhshsvrrairpredkeell 60
 QY 61 MHNKLRGVQPOASNMEYMTWDELKSAAMASOCIEHGPTSLVSIIGNLGAHWGR 120
 DB 61 mhnklrgvqpoasnmeymtwdelekasaawaasqciwebhptslvsiignlgahwgr 120
 QY 121 YRSPGFHVSQWYDEVKDYTPRPSECNPCPCRCGPMCTHTQIYMATNTKIGCAVNTC 180
 DB 121 yrspgfhwswydevkdytrpsecpncpcrcgpmcthytqlwatkngicavntc 180
 QY 181 RKMATWGEWENAVYFVCNYSFKGNMIGEPYKNGRPSCEPSPYSGSCNNLCYREETY 240

Db 181 rkmrtwgevenavayfvcnyspknwlgaeaykngdrpcsecpysgscnmlcyreety 240
 Qy 241 TPKEPDEDMNEVETAPIPEDENHVMLOPRVMPRTKPKTSANVMYQVRCOTKMKDRCKG 300
 Db 241 tpxpeldemeveletapipreenhvalqprvmpkpktsavymtqvrcdtkmkdrckg 300
 Qy 301 STCRNRYOCPCAGCLNHRKAKIFGSLFESSSSICRAAIHYGLIDDKGGLVDITRNGKVPFF 360
 Db 301 stctnryqcpagclnhkakiifgslfessssicraaihygliddkgglvdlitngkvpff 360
 Qy 361 KSERHGVOSLSKTKPSSFWKVKYODLDCYTTVAOLCPFEKPAHPCRIHCPAHCKDE 420
 Db 361 kserhgvoslsktpssfwkvykqldcyltvaqlcpfekpachpcrihpnachkd 420
 Qy 421 PSYMAVPGVGNITADTSSTCKTAVHAGVISNESGVDVMPVKKTKYVGSILRNGVOS 480
 Db 421 psywavpgvgnitadtsstcktavhagvisnesgvdvmpvdkkkyvgsilrnyvgs 480
 Qy 481 LGTPRDGKAFFRFAVRQ 497
 Db 481 lgtprdgkafrifavrq 497

RESULT 3
 ID AAE10618 standard; Protein; 498 AA.
 AC AAE10618;
 DT 10-DEC-2001 (first entry)

Human novel trypsin inhibitor-like protein, NOV-4d.
 Human NOV-X protein; KIAA1233-like protein; STE20-like protein; tumour;
 trypsin inhibitor-like protein; gene therapy; haematopoietic; illness;
 immunological disorder; neurodegenerative disorder; Alzheimer's disease;
 Parkinson's disease; immunomodulatory; pharmacogenomic; haemostatic;
 human immunodeficiency virus; HIV; fertility disorder; neuroprotective;
 cystostatic; neurotropic; anti-fertility; cancer; NOV-4d protein.
 Homo sapiens.
 FH Key Location/Qualifiers
 FT Domain 196..207
 FT /label= SCP_domain
 PN MO200162928-A2.
 PD 30-AUG-2001.
 PF 26-FEB-2001; 2001WO-0506151.
 PR 25-FEB-2000; 2000US-0184951.
 PR 28-FEB-2000; 2000US-018548.
 PR 01-MAR-2000; 2000US-018548.
 PR 18-APR-2000; 2000US-0197423.
 PR 27-APR-2000; 2000US-0199957.
 PR 23-FEB-2001; 2001US-0789390.
 PA (CORA-) CURAGEN CORP.
 PI Vernet CAM, Fernandes E, Shankets RA, Macdougall J, Spaderna SK,
 DR WPI; 2001-582051/65.
 DR N-PSDB; AAD17766.
 DR XX
 PT New isolated KIAA1233-like, STE20-like, or trypsin inhibitor-like
 PT polypeptide for diagnosing and treating pathological disorders, such as
 PS Parkinson's disease and for use in pharmacogenomics
 XX Claim 1; Page 88; 189pp; English.

CC The invention relates to novel human polypeptides referred as NOV-X
 CC and their corresponding nucleic acid sequences. NOV-X collectively
 CC include NOV-1, NOV-2a and NOV-2b which are novel KIAA1233-like
 CC polypeptides, NOV-3a, NOV-3b, NOV-3c and NOV-3d which are novel
 CC STE20-like polypeptides and NOV-4a, NOV-4b, NOV-4c, NOV-4d and NOV-4e
 CC which are novel trypsin inhibitor-like polypeptides. NOV-4d and NOV-4e
 CC identify a potential therapeutic agent that can modulate its activity
 CC and can be used for treating a pathology related to aberrant expression
 CC or aberrant physiological interactions of NOV-X. NOV-X or its DNA is
 CC used to determine the presence or predisposition to a disease associated
 CC with altered levels of NOV-X. NOV-X, its DNA and its antibody are used
 CC to treat or prevent a pathology associated with NOV-X. The pathological
 CC states that can be treated or prevented are haematopoietic, cancer,
 CC immunological, tumour, neurodegenerative (e.g. Alzheimer's and
 CC Parkinson's disease), human immunodeficiency virus (HIV) illness and
 CC fertility disorders. NOV-X and its DNA are used in pharmacogenomics for
 CC predictive medicine. NOV-X DNA is used in gene therapy. The present
 CC sequence is human novel trypsin inhibitor-like protein, NOV-4d.
 SQ Sequence 498 AA:

Query Match 99.2%; Score 2736.5; DB 22; Length 498;
 Best Local Similarity 99.6%; Pred. No. 3.1e-226;
 Matches 496; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Qy 1 MSCVIGGVIPILGLIFVCSOGYILPNVTLLESLSKYONESHVRRAIPREDKEEIL 60
 Db 1 mscvlgvvpilglifvcsqgyllpnvtlleelskqhneshvrratpredkeell 60
 Qy 61 MLHNKLRGVQVPOASNMENMTWDELEKSAAMAASQCTIEHPPTSLVSIQNTLGHMG- 119
 Db 61 mlhnklrgvqvpoasnmewmtwdeleksaamaasqctiewhptslvsiqnlghmwr 120
 Qy 120 RRSRGFRVQSYDENVNITYPSECNPCWPCERCSGPCCTHYTOIVATTKIGCAVNT 179
 Db 121 rrsrgfrvqsydyenvnitypsecnpcwpcercsGPCCTHYTOIVATTKIGCAVNT 179
 Qy 180 CRKMTWGEVWENAVYFVCNYSKGNWIGEPYKNGRPSCEPSPYSGSCRNMLCYREET 239
 Db 181 crkmtwgevenavayfvcnyspknwlgaeaykngdrpcsecpysgscnmlcyreet 240
 Qy 240 YTPKEPDEDMNEVETAPIPEDENHVMLOPRVMPRTKPKTSANVMYQVRCOTKMKDRCK 299
 Db 241 ytpkeldemeveletapipreenhvalqprvmpkpktsavymtqvrcdtkmkdrck 300
 Qy 300 GSTCRNRYOCPCAGCLNHRKAKIFGSLFESSSSICRAAIHYGLIDDKGGLVDITRNGKVPFF 359
 Db 301 gstcnryqcpagclnhkakiifgslfessssicraaihygliddkgglvdlitngkvpff 360
 Qy 360 VKSERHGVOSLSKTKPSSFWKVKYODLDCYTTVAOLCPFEKPAHPCRIHCPAHCKD 419
 Db 361 vkserhgvoslsktpssfwkvykqldcyltvaqlcpfekpachpcrihpnachkd 420
 Qy 420 EPSTYMAVPGVGNITADTSSTCKTAVHAGVISNESGVDVMPVKKTKYVGSILRNGVOS 479
 Db 421 epsywavpgvgnitadtsstcktavhagvisnesgvdvmpvdkkkyvgsilrnyvgs 480
 Qy 480 LGTPRDGKAFFRFAVRQ 497
 Db 481 lgtprdgkafrifavrq 498

RESULT 4
 ID AAE10619 standard; Protein; 501 AA.
 AC AAE10619;
 DT 10-DEC-2001 (first entry)
 DT Human novel trypsin inhibitor-like protein, NOV-4e.
 DE XX

BASE COUNT 695 a 851 c 851 g 691 t
ORIGIN

Db	Accession	Source	Gene	Protein	Length	Score	E-value
Db	912	ACTCCAAAACCTGAAACGGACCGAGATGATGAGTGGAAACGGCTCCATCTCTGAAGA			971		
Qy	781	aaccatggttggctcccaaccgaggtgatgagaccaccaagcccaagaanaacccctgcg			840		
Db	972	AACCATGTTGGCTCCAAACCGAGGGTGAAGAACCCACCAAGCCCAAGAAACCTCTGGG			1031		
Qy	841	gtcaactacatcaccaagtcgtcagatgtgacacccaatgaagaagagacagtgcaagag			900		
Db	1032	GTCACACTACATACACCCAACTCTCATGATGTGACACCCAAAGTGAAGGACAGGTGCAAAAGG			1091		
Qy	901	tcacagctgaacaggtacacagtgcccagagagctgctgaacacacaagcgagaatcttt			960		
Db	1092	TCCACGCTGTAAACAGGTACACAGTCCACAGAGGCTGCTGCTGAACCAAGCGAAGATCTTT			1151		
Qy	961	ggaagtcgttcataagaagctcgtctgacatatgccgcgcgcgcacatcactaaggatc			1020		
Db	1152	GGAATCTGTTCTATGAAAGCTGCTGTGACATATCCCGGCCCATCCATCACAGGAGATC			1211		
Qy	1021	ctgagtcgaagagggagcctggtgatataccacaagaaacgaggaagtcacctcttcgtg			1080		
Db	1212	CTGGATCTCAAGAGGAGGCGCTGCTGATATCACACGAAGAGGAGATCCCTCTTCTCGTG			1271		
Qy	1081	aagctctgagagacagcggtgtcagatccctcagcaaa			1116		
Db	1272	AAGTCTGAGAACACAGCGCTGACATCCCTCAGAGTAA			1307		
RESULT	14						
LOCUS	AF109674						
DEFINITION	Rattus norvegicus late gestation lung protein 1 (Lg11) mRNA,						
VERSION	AF109674						
KEYWORDS	AF109674.1 GI:4324681						
ORGANISM	Norway rat.						
REFERENCE	Rattus norvegicus						
AUTHORS	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
TITLE	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;						
JOURNAL	Rattus.						
MEDLINE	1 (bases 1 to 3052)						
PubMed	Kaplan, F., Ledoux, P., Kassamali, F.O., Gagnon, S., Post, M.,						
REFERENCE	Koehler, D., Delmink, J., and Swezey, N.B.						
AUTHORS	A novel developmentally regulated gene in lung mesenchyme: homology						
TITLE	to a tumor-derived trypsin inhibitor						
JOURNAL	Am. J. Physiol. 276 (6), L1027-L1036 (1999)						
FEATURES							
source	Submitted (26-NOV-1998) Human Genetics, McGill University- Montreal						
	Children's Hospital Research Institute, 2300 Tupper Street,						
	Montreal, QC H3H 1P3, Canada						
	Location/Qualifiers						
	1. 3052						
	/organism="Rattus norvegicus"						
	/strain="Wistar"						
	/db_xref="taxon:10116"						
	/cell_type="Fibroblast"						
	/tissue_type="lung"						
	/dev_stage="20 day old fetus"						
	1. 3052						
	/gene="Lg11"						
	/note="horizontally and developmentally regulated"						
	265..831						
	/gene="Lg11"						
	/note="expressed in late gestation lung mesenchymal cells						
	in rat, mouse, and human; similar to tumor-derived trypsin						
	inhibitor."						
	/codon_start=1						
	/product="late gestation lung protein 1"						

Query Match	32.5%	Score 484.4;	DB 6;	Length 2664;
Best Local Similarity	59.8%	Pred. No. 2.6e-105;		
Matches 875;	Conservative	0;	Mismatches 566;	Indels 21;
			Gaps	3;

Db	147	TCGCTTTCATGAGCTAAGACAACTTCACGACATGCGTGGTTCCCAATGCCACTTTATTGGAGA	206
QY	98	agctgctcaagaataaccagacaagaagctcaactcccggtgcg-----	144
Db	207	AACCTTTTGGAAAAATTAACATGGATAGAGATGGTGTGATGGTGGATAGACCAACAACGAGGGA	266
QY	145	--agagccatcccccaggagagacagaaggagatcccacatgcgcgcacacaagaactccgg	202
Db	267	AAAGGCCATACAGACACATACATGCAAGATATTTGGACCTTCATAAATTAATTCGAA	326
QY	203	gccaggtgcagcctcaggtccacacatggagtacatgcactggaactggtgagaaactgcgag	262
Db	327	GTCAGGTGTATCCAAACAGCCTCTTAATGTAGATATATGACATGAGTATGTAGACTGGAAA	386
QY	263	atgtctgcgcagcggtggggccaagtcaatgcatctcgggagcagcgcccaacagctcgtcg	322
Db	387	GATTCGCAGAACTCTTGCGCTGGAAAGTTGCTGTGGGAAACATGTGACCTGCACAGCTTCTTC	446
QY	323	tgatcatcgggcagaaacctggcgctcaactcagcgagagtagatcgctccgggttccatg	382
Db	447	CATCAATTTGGACAGAAATTTGGAGCACACAGGGGAATATAGCCCCCGACGCTTTCATG	506
QY	383	tgacgtcttgatgaacgaggttgaaagtaacatacaccatacccgagcgagtgcnaac	442
Db	507	TACATGCTGGTATGAATGAAGTGAACAACTTAGCTACCATATGAAACATGATGCACAC	566
QY	443	ccgtgttccggagaagttgtctcggggcctatgycagcgcactacacagataagtttgg	502
Db	567	CATATGTTCATTCAGGTGTGTTCGGCCTTATGTACACTTATPACACAGCTGTGTGGG	626
QY	503	ccaccaccaaaatctggttgtgtgtgtaacaccttcgcggaagatgcatctgtctggag	562
Db	627	CAATCTACTAACAGAACTCGTTGTGCCATTAATTGTGTATACATGAAACATCTGGGGC	686
QY	563	aagttggaggaacgpgttacttgttctgcgaattatctccaaagggaaactgattg	622
Db	687	AGATATGGCCCAAGCTGTCTTACCTGGTGTGCATTAATCCCAAGGGAAACCTGGTGG	746
QY	623	gagaagccccctacaagaatgycgggcctgtctagtgcgcacccagctatgtgagaa	682
Db	747	GCAATGCCCTTACAAACATGGGGCGGCGCGTGTCTGTGCCACCTAGTTTGGAGGGG	806
QY	683	gtgcgcggaacaactgtgttaccgaggaagaaactacaactccaaaacttgaacgagcg	742
Db	807	GCTGTAGAGAAATTCGTGTGTACAAAGAGGTCACAGACAGATTAATCCCTCCGAGAG	866
QY	743	agatgaatgagtgagaaagcgctcccatctccagaagaaacatgtttggtcccaacga	802
Db	867	AGGAAACAAT---GAATATGAAACGACACACTATCACAACTCATGACACCCATGCCGA	923
QY	803	gggtgtgagaccaccaagcccaagaanaactctgtcgttcaactacatgaccaagtcg	862
Db	924	CAAGATCAGATGATATAGACAGAAATGAAGATCATTAAGCCACAGCAAAATGTCCAATTG	983
QY	863	tcaagtgtgaaccaccaagatgaaaggaagatgtgaagaaggttccacgtgttaacagttcac	922
Db	984	TTTCTTGTGAAGTAAAGATTAAAGATCACTGCAAAAGGAACAACTGCACATAGTACGAT	1043
QY	923	gcccaagcaggtcgtcgtgaacccaagaagcgaaagatcttggaaatctgtgtcatgaaagt	982
Db	1044	GTCCTCTGCTGGCTGTTGGATGTAAGCTAAAGCTAAAGTTATGGCGATTAATATATGAAATGC	1103
QY	983	cgtctgacatagccgcgcgcacatccactaaggtatccctgtgatacaagaagagcgctg	1042
Db	1104	AATTCAGCATCTGTAGAGCTGCAATTCATTATAGTATATAGACAAATATGCTGGCTGGG	1163
QY	1043	tggataccaccaggaacgggaaggttccctcttctgtgaagtctgagaaacacggcggtgc	1102
Db	1164	TGATATCTACTGAGACAAGGAACAAAGCATATATTTTCATCAAGTCCAAATGAAATGATTC	1223
QY	1103	agtcctcaggaataataaaccttccagctaatcatcattcattggtgtcaaaagtgaagtgcg	1162
Db	1224	AAACATTTGGCAAAATATACGTCTGTATTTCTTACACAGCTCTTAAGATACATTCG	1283

Tue May 28 16:10:03 2002

copy 2 (p. 2 + 3)

us-09-667-380a-2.rag

PD 9/24/1999 only 8 in ok (data), copy 6-8 Page 1

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 25, 2002, 13:22:04 ; Search time 76.94 Seconds

(without alignments)
717.490 Million cell updates/sec

Title: US-09-667-380A-2

Perfect score: 2759

Sequence: 1 MSCVGGYIPGLFLVCGS.....SLSLGPDRGKAFRIFAVRQ 497

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A: Geneseq_032802: *
1: /SIDS1/gcgcdata/geneseq/geneseq-emb1/AA1980.DAT: *
2: /SIDS1/gcgcdata/geneseq/geneseq-emb1/AA1981.DAT: *
3: /SIDS1/gcgcdata/geneseq/geneseq-emb1/AA1982.DAT: *
4: /SIDS1/gcgcdata/geneseq/geneseq-emb1/AA1983.DAT: *
5: /SIDS1/gcgcdata/geneseq/geneseq-emb1/AA1984.DAT: *
6: /SIDS1/gcgcdata/geneseq/geneseq-emb1/AA1985.DAT: *
7: /SIDS1/gcgcdata/geneseq/geneseq-emb1/AA1986.DAT: *
8: /SIDS1/gcgcdata/geneseq/geneseq-emb1/AA1987.DAT: *
9: /SIDS1/gcgcdata/geneseq/geneseq-emb1/AA1988.DAT: *
10: /SIDS1/gcgcdata/geneseq/geneseq-emb1/AA1989.DAT: *
11: /SIDS1/gcgcdata/geneseq/geneseq-emb1/AA1990.DAT: *
12: /SIDS1/gcgcdata/geneseq/geneseq-emb1/AA1991.DAT: *
13: /SIDS1/gcgcdata/geneseq/geneseq-emb1/AA1992.DAT: *
14: /SIDS1/gcgcdata/geneseq/geneseq-emb1/AA1993.DAT: *
15: /SIDS1/gcgcdata/geneseq/geneseq-emb1/AA1994.DAT: *
16: /SIDS1/gcgcdata/geneseq/geneseq-emb1/AA1995.DAT: *
17: /SIDS1/gcgcdata/geneseq/geneseq-emb1/AA1996.DAT: *
18: /SIDS1/gcgcdata/geneseq/geneseq-emb1/AA1997.DAT: *
19: /SIDS1/gcgcdata/geneseq/geneseq-emb1/AA1998.DAT: *
20: /SIDS1/gcgcdata/geneseq/geneseq-emb1/AA1999.DAT: *
21: /SIDS1/gcgcdata/geneseq/geneseq-emb1/AA2000.DAT: *
22: /SIDS1/gcgcdata/geneseq/geneseq-emb1/AA2001.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2759	100.0	497	22	AAAB74446
2	2743	99.4	497	22	AAAE10616
3	2736.5	99.2	498	22	AAAE10618
4	2721	98.6	501	22	AAAE10619
5	2595	94.1	477	22	AAAE10617
6	2140	77.6	383	22	AAAE10615
7	2071	75.1	371	22	AAAB3979
8	1640.5	59.5	300	20	AAAY1738
9	1640.5	59.5	500	21	AAAB4294
10	1640.5	59.5	500	22	AAU29058
11	1407.5	51.0	522	22	AAAM1693

ALIGNMENTS

RESULT 1	AAAB74446	standard; Protein; 497 AA.
AC	AAAB74446;	
DT	29-MAY-2001	(first entry)
XX	Human	protease-inhibitor like protein.
DE	Human	protease-inhibitor like protein.
XX	Human	protease-inhibitor like protein.
KW	Human	protease-inhibitor like protein; NRP; novel human protein;
KW	trypsin inhibitor; gene therapy; drug screening; disease diagnosis;	
XX	signal transduction.	
OS	Homo sapiens.	
XX	WO200121651-A2.	
PN	29-MAR-2001.	
XX	22-SEP-2000; 2000WO-US26048.	
PF	24-SEP-1999; 99US-0156101.	
XX	(LEXI-) LEXICON GENETICS INC.	
PA	Donoho G, Turner CA, Wattler F, Nehls M, Friedrich G;	
XX	Zambrowicz B, Sands AT;	
PI	WPI: 2001-266064/27.	
DR	N-ESDB; AAF77686; AAF77687.	
DR	Novel isolated human protease inhibitor-like polynucleotide useful in	
XX	therapeutic, diagnostic and pharmacogenic applications	
PT		

XX Claim 2; Page 27-28; 29pp; English.

XX The present invention provides the protein and coding sequences of the
 CC novel human protease-inhibitor like protein NHP. This shows homology to
 CC mammalian trypsin inhibitors. The sequences are useful in disease
 CC diagnosis and treatment, particularly of diseases associated with signal
 CC transduction. The present sequence is the NHP protein.

XX Sequence 497 AA:

Query Match 100.0%; Score 2759; DB 22; Length 497;
 Best Local Similarity 100.0%; Pred. No. 3,66-228;
 Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MSCVIGVPLGLFLVCGSGGYLLPNVTLLLEELLSKYQHNSHSRRRAIPREDKEEL 60
 1 mscvlgvplglflvsgsggyllpnvtllleellskvqhnsrsvrralpredkeell 60
 61 MLHNKLRGOVPOASNMETMTWDELEKSAAMASQCIWEHGPSTLSVIGNLGAHNGR 120
 61 mlhnklrgovpogasnmetymtwdeleksaasqciwehgpstlsvlgnglgaahngr 120
 121 YRSPGFHVSQWYDEVKDYTPYPSECPNCPERCSPMCTHTQTQIWMATNKGAVNTC 180
 121 yrspgfhwswydevkdyltppsecnpwpcercspmcthtqtlwatkngavntc 180
 181 RKMTWGEWENAVYFVCNYSKGNWIGEADPYKNGRPSCEPSTGGSCRNNLCYREETY 240
 181 rkmtwgewenavyfvcnyskgnwigeadpykngprscpspgsgrnnlcyreety 240
 241 TPRPTEEMNEVEETAPRPEENHWLLOPRVMPRTKRTSAVYMTQVYRCDKMDRCKG 300
 241 tprptemneveetaprpeenhwlloprvmptrtkrtsavymtqvyrckmdrckg 300
 301 STCRNYOCPAGCLNHNKAKIFGSLFEYSSSSICRAAIHYGIIDDKGLVDITRNGKVPFV 360
 301 stcrnyocpagclnhnkakifgslfeyssssicraaihygiiddkglvdltrngkvpfv 360
 361 KSEHGVQSLSKYPSSFFWYSKRVQDLDCYTAOLCPFEKPTHCRIHCPAKCKDE 420
 361 ksehgvqslskypssffwyskrvqdlscytaolcpefkpthcricpachckde 420
 421 PSYAPFEGTITADTSSICKTAHAGVINSNCGDVPDMPYDKKRTYVGLRNGVQSES 480
 421 psyapfegtitadtssicktahagvinsnsgdvpdmpydkkrtvyglsrngvses 480
 481 LGTRPDGKAERIPAVRQ 497
 481 lgtprdgkafrifavrq 497

RESULT 2

AAE10616 standard; Protein: 497 AA.

AAE10616:

10-DEC-2001 (first entry)

Human novel trypsin inhibitor-like protein, NOV-4b.

Human; NOV-X protein; KIA1233-like protein; STE20-like protein; tumour;
 trypsin inhibitor-like protein; gene therapy; haematopoietic; illness;
 immunological disorder; neurodegenerative disorder; Alzheimer's disease;
 Parkinson's disease; immunomodulatory; pharmacogenomic; haemostatic;
 human immunodeficiency virus; HIV; fertility disorder; neuroprotective;
 cytosolic; nootropic; anti-fertility; cancer; NOV-4b protein.

Homo sapiens.
 Location/Qualifiers

SEA ID NO: 2
 AE NO: AAE10616
 Database: A-Geneseg-032802

Peptide 1..22
 /label= signal_peptide
 Protein 23..497
 /note= "Human mature novel trypsin inhibitor-like
 protein, NOV-4b"
 Domain 195...206
 /label= SCP_domain

MO200162928-A2.

30-AUG-2001.

26-FEB-2001; 2001WO-US06151.

25-FEB-2000; 2000US-0184951.

28-FEB-2000; 2000US-0185548.

01-MAR-2000; 2000US-0185967.

18-APR-2000; 2000US-0197723.

27-APR-2000; 2000US-0199957.

23-FEB-2001; 2001US-0789390.

(CURA-) CURAGEN CORP.

Vernet CAM, Fernandes E, Shinkets RA, Macdougall J, Spaderna SK;

WPI; 2001-582051/65.

N-PSDB: AAD17764.

New isolated KIA1233-like, STE20-like, or trypsin inhibitor-like

polypeptide for diagnosing and treating pathological disorders, such as

Parkinson's disease and for use in pharmacogenomics

Claim 1; Page 83; 189pp; English.

The invention relates to novel human polypeptides referred as NOV-X
 and their corresponding nucleic acid sequences. NOV-X collectively
 include NOV-1, NOV-2a and NOV-2b which are novel KIA1233-like
 polypeptides, NOV-3a, NOV-3b, NOV-3c and NOV-3d which are novel
 STE20-like polypeptides and NOV-4a, NOV-4b, NOV-4c, NOV-4d and NOV-4e
 which are novel trypsin inhibitor-like polypeptides. NOV-X is used to
 identify a potential therapeutic agent that can modulate its activity
 or can be used for treating a pathology related to aberrant expression
 or aberrant physiological interactions of NOV-X. NOV-X or its DNA is
 used to determine the presence or predisposition to a disease associated
 with altered levels of NOV-X. NOV-X, its DNA and its antibody are used
 to treat or prevent a pathology associated with NOV-X. The pathological
 states that can be treated or prevented are haematopoietic, cancer,
 immunological, tumour, neurodegenerative (e.g. Alzheimer's and
 Parkinson's disease), human immunodeficiency virus (HIV) illness and
 fertility disorders. NOV-X and its DNA are used in pharmacogenomics for
 predictive medicine. NOV-X DNA is used in gene therapy. The present
 sequence is human novel trypsin inhibitor-like protein, NOV-4b.

Sequence 497 AA:

Query Match 99.4%; Score 2743; DB 22; Length 497;

Best Local Similarity 99.6%; Pred. No. 8,66-227;

Matches 495; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 MSCVIGVPLGLFLVCGSGGYLLPNVTLLLEELLSKYQHNSHSRRRAIPREDKEEL 60
 1 mscvlgvplglflvsgsggyllpnvtllleellskvqhnsrsvrralpredkeell 60
 61 MLHNKLRGOVPOASNMETMTWDELEKSAAMASQCIWEHGPSTLSVIGNLGAHNGR 120
 61 mlhnklrgovpogasnmetymtwdeleksaasqciwehgpstlsvlgnglgaahngr 120
 121 YRSPGFHVSQWYDEVKDYTPYPSECPNCPERCSPMCTHTQTQIWMATNKGAVNTC 180
 121 yrspgfhwswydevkdyltppsecnpwpcercspmcthtqtlwatkngavntc 180
 181 RKMTWGEWENAVYFVCNYSKGNWIGEADPYKNGRPSCEPSTGGSCRNNLCYREETY 240

```

Db 181 fkmfhwgevenavayfvcnyspkgnwlgaeaykngprscscppsygscrmncyreety 240
QY 241 TPKEPTDENMEVEETAPLPEENHVMLOPRVMPRTKPKKTSAVNMTQVVRCDTKMKDRCKG 300
Db 241 tprkpetdemnevetapripeenhvwlqprvmrptkpkktksavnymtqvvrcdtkmkdrckg 300
QY 301 STCRNRYQCPAGCLNHRKAKIFGSLFYESSSSICRAAIHVGILDDKGGVLDITRNGKVPFFV 360
Db 301 stcrnryqcpagclnhkakifgslfyessssicraahvgilddkgylditrngkvpffv 360
QY 361 KSEHNGVOSLSKYKPPSSFMVSKVQVODLDCTYTTVAOLCPFEKPATHCPRTHCAHCKDE 420
Db 361 kserhngvqslskypssfmvskkvqvdldctyttvaqlcpfekpathcprthcapckde 420
QY 421 PSYAPVFGNTIYADTSSICKTAHVAGVISNESGGVDVMPVDKRTYVGSILRNGVQSES 480
Db 421 psyapvfgntiyadtssicktahvagvisnesgvdvmpvdkkkyvgsilrngvses 480
QY 481 LGTPRDGKAFRIFAARQ 497
Db 481 lgtprdgkafrifavrg 497

RESULT 3
AAE10618
ID AAE10618 standard; Protein; 498 AA.
AC AAE10618;
DE 10-DEC-2001 (first entry)
XX Human novel trypsin inhibitor-like protein, NOV-4d.
XX Human; NOV-X protein; KIAA1233-like protein; STE20-like protein; tumour;
XX trypsin inhibitor-like protein; gene therapy; haematopoietic; illness;
XX immunological disorder; neurodegenerative disorder; Alzheimer's disease;
XX Parkinson's disease; immunomodulatory; pharmacogenomic; haemostatic;
XX human immunodeficiency virus; HIV; fertility disorder; neutroprotective;
XX cytosolic; neutroptic; anti-fertility; cancer; NOV-4d protein.
XX Homo sapiens.
XX OS
XX FH Key Location/Qualifiers
XX FT Domain 196..207
XX FT /label= SCP_domain
XX PN WO200162928-A2.
XX PD 30-AUG-2001.
XX PF 26-FEB-2001; 2001WO-US06151.
XX PR 25-FEB-2000; 2000US-0184951.
XX PR 28-FEB-2000; 2000US-0185448.
XX PR 01-MAR-2000; 2000US-0185967.
XX PR 18-APR-2000; 2000US-0197723.
XX PR 27-APR-2000; 2000US-0199957.
XX PR 23-FEB-2001; 2001US-0789390.
XX PA (CUBA-) CUBAGEN CORP.
XX PI Vernet CAM, Fernandes E, Shinkets RA, Macdougall J, Spaderna SK;
XX DR WPI: 2001-582051/65.
XX DR N-PSDB: AAD17766.
XX PT New isolated KIAA1233-like, STE20-like, or trypsin inhibitor-like
XX PT polypeptide for diagnosing and treating pathological disorders, such as
XX PT Parkinson's disease and for use in pharmacogenomics -
XX PS Claim 1; Page 88; 189pp; English.
XX XX

```

```

CC The invention relates to novel human polypeptides referred as NOV-X
CC and their corresponding nucleic acid sequences. NOV-X collectively
CC include NOV-1, NOV-2a and NOV-2b which are novel KIAA1233-like
CC polypeptides, NOV-3a, NOV-3b, NOV-3c and NOV-3d which are novel
CC STE20-like polypeptides and NOV-4a, NOV-4b, NOV-4c, NOV-4d and NOV-4e
CC which are novel trypsin inhibitor-like polypeptides. NOV-X is used to
CC identify a potential therapeutic agent that can modulate its activity
CC and can be used for treating a pathology related to aberrant expression
CC or aberrant physiological interactions of NOV-X. NOV-X or its DNA is
CC used to determine the presence or predisposition to a disease associated
CC with altered levels of NOV-X. NOV-X, its DNA and its antibody are used
CC to treat or prevent a pathology associated with NOV-X. The pathological
CC states that can be treated or prevented are haematopoietic, cancer,
CC immunological, tumour, neurodegenerative (e.g. Alzheimer's and
CC Parkinson's disease), human immunodeficiency virus (HIV) illness and
CC fertility disorders. NOV-X and its DNA are used in pharmacogenomics for
CC predictive medicine. NOV-X DNA is used in gene therapy. The present
CC sequence is human novel trypsin inhibitor-like protein, NOV-4d.
XX
XX Sequence 498 AA;
SQ

```

```

Query Match 99.2%; Score 2736.5; DB 22; Length 498;
Best Local Similarity 99.6%; Pred. No. 3; Le-226;
Matches 496; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

```

```

QY 1 MSCVLAGVPIPLGLFLVNGSOGYLLPNVTLLELLSKYQHNHSRRVRAIPREDKEITL 60
Db 1 mscvlagvpiplgllflvngsgyllpnvttllellskymeshsrvtalpredkeil 60
QY 61 MLHNKLRGQVQPOASNNEMYTWDELKSAAMASOCITWEGPTSLVSIQNLGAHMG- 119
Db 61 mlhnklrgvqvgqasnnemymtwdelksaawasqciwhegptslsvsiqnlghabgr 120
QY 120 RYRSPGFVQSVWDEVDYTPYPSSECPWCPREGSGPMCHRYQIYVATTNKGCAVNT 179
Db 121 ryrspgfhwgwydevdytlypypsecpwcpercsgpmchycqivwatnkgicavnt 180
QY 180 CRKMTWGEWENAVFYVCNTPSKGNWIGEAPEYKNGRSCSPSYGSCRNNLCYREET 239
Db 181 crkmtwgevenavayfvcnyspkgnwlgaeaykngprscscppsygscrmncyreety 240
QY 240 YTPKEPTDENMEVEETAPLPEENHVMLOPRVMPRTKPKKTSAVNMTQVVRCDTKMKDRCK 299
Db 241 ytpkpetdemnevetapripeenhvwlqprvmrptkpkktksavnymtqvvrcdtkmkdrck 300
QY 300 GSTCNRVQCPAGCLNHRKAKIFGSLFYESSSSICRAAIHVGILDDKGGVLDITRNGKVPFF 359
Db 301 gstcnryqcpagclnhkakifgslfyessssicraahvgilddkgylditrngkvpff 360
QY 360 VKSEHNGVOSLSKYKPPSSFMVSKVQVODLDCTYTTVAOLCPFEKPATHCPRTHCAHCKD 419
Db 361 vkserhngvqslskypssfmvskkvqvdldctyttvaqlcpfekpathcprthcapckd 420
QY 420 EPSYAPVFGNTIYADTSSICKTAHVAGVISNESGGVDVMPVDKRTYVGSILRNGVQSES 479
Db 421 epsyapvfgntiyadtssicktahvagvisnesgvdvmpvdkkkyvgsilrngvse 480
QY 480 SLGTPRDGKAFRIFAARQ 497
Db 481 slgtprdgkafrifavrg 498

RESULT 4
AAE10619
ID AAE10619 standard; Protein; 501 AA.
AC AAE10619;
DE 10-DEC-2001 (first entry)
XX Human novel trypsin inhibitor-like protein, NOV-4e.
XX XX

```

KM Human: NOV-X protein: KIAA1233-like protein; STE20-like protein; tumour;
 KM trypsin inhibitor-like protein; gene therapy; haematopoietic; illness;
 KM immunological disorder; neurodegenerative disorder; Alzheimer's disease;
 KM Parkinson's disease; immunomodulatory; pharmacogenomic; haemostatic;
 KM human immunodeficiency virus; HIV; fertility disorder; neuroprotective;
 KM cytosolic; neotropic; anti-fertility; cancer; NOV-4c protein.
 OS Homo sapiens.
 XX
 PH Key Location/Qualifiers
 FT Peptide 1..22
 FT /label= Signal_peptide
 FT Protein 23..501
 FT /note= "Human mature novel trypsin inhibitor-like
 protein, NOV-4c"
 XX
 PN WO200162928-A2.
 PD 30-AUG-2001.
 XX
 PF 26-FEB-2001; 2001WO-US06151.
 XX
 PR 25-FEB-2000; 2000US-0184951.
 PR 28-FEB-2000; 2000US-0185348.
 PR 01-MAR-2000; 2000US-0185967.
 PR 18-APR-2000; 2000US-0197723.
 PR 27-APR-2000; 2000US-0199957.
 PR 23-FEB-2001; 2001US-0789390.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Vernet CAM, Fernandes E, Shimkets RA, Macdougall J, Spaderna SK;
 XX
 DR WPI: 2001-582051/65.
 DR N-PSDB: AAD17767.
 PT New isolated KIAA1233-like, STE20-like, or trypsin inhibitor-like
 PT polypeptide for diagnosing and treating pathological disorders, such as
 PT Parkinson's disease and for use in pharmacogenomics -
 PS Claim 1; Page 91; 189pp; English.
 XX
 XX The invention relates to novel human polypeptides referred to as NOV-X
 CC and their corresponding nucleic acid sequences. NOV-X collectively
 CC include NOV-1, NOV-2a and NOV-2b which are novel KIAA1233-like
 CC polypeptides, NOV-3a, NOV-3b, NOV-3c and NOV-3d which are novel
 CC STE20-like polypeptides and NOV-4a, NOV-4b, NOV-4c, NOV-4d and NOV-4e
 CC which are novel trypsin inhibitor-like polypeptides. NOV-X is used to
 CC identify a potential therapeutic agent that can modulate its activity
 CC and can be used for treating a pathology related to aberrant expression
 CC or aberrant physiological interactions of NOV-X. NOV-X or its DNA is
 CC used to determine the presence or predisposition to a disease associated
 CC with altered levels of NOV-X. NOV-X, its DNA and its antibody are used
 CC to treat or prevent a pathology associated with NOV-X. The pathological
 CC states that can be treated or prevented are haematopoietic, cancer,
 CC immunological, tumour, neurodegenerative (e.g. Alzheimer's and
 CC Parkinson's disease), human immunodeficiency virus (HIV) illness and
 CC fertility disorders. NOV-X and its DNA are used in pharmacogenomics for
 CC predictive medicine. NOV-X DNA is used in gene therapy. The present
 CC sequence is human novel trypsin inhibitor-like protein, NOV-4c.
 XX
 SQ Sequence 501 AA:
 Query Match 98.6%; Score 2721; DB 22; Length 501;
 Best Local Similarity 98.8%; Pred. No. 6,7e-225;
 Matches 495; Conservative 0; Mismatches 2; Indels 4; Gaps 2;

DB 61 mlnhklrgvqpsasmeymtwddelaksaawasqclwelpqllvsigqlghvgr 120
 QY 121 YRSRGFVQSVYDVKQYTYTPYSECNWCPERGPMCTHY---TOYVMAATNKIGAV 177
 DB 121 YRSPGfhwqsvydwkdylypypsecpwpccpsgmclhlyqlydwatlnkigav 180
 QY 178 NTCRKMTVWGEVWENAVYFVNCYSPK-GNWIGEARPYKNGRPCSECPYSGSCRNMLCYR 236
 DB 181 ntcckmtvgevenavlyfvcnypkrgnwageapykngrpccseppayggsccrnmlcyr 240
 QY 237 EETYPKPEIDENNEVEETAPLPEENHWLQPRWAPPTPKTSAVNWATQVYRCOTKAKD 296
 DB 241 eelypkpeldemneveletapieehwvldprvmrptkpktsavnymqvvrcdtkmkd 300
 QY 297 RCKGTCNRVCCPAGCCLNHKAKIFGSLFEYSSSSICRAIHGILDDGGVLDITRNKGV 356
 DB 301 rckgstcnryqcpagclnhkakifgslfeyssscraahgildggylditrnkgy 360
 QY 357 PFVKSERHGVQSLSKKPPSSFMVSKVYKVDLDCTTVAQLCPPEKPAHCPRIHCPAH 416
 DB 361 pfvksrhwqsvlskppssfmvskvkvqldcyltvaqlcpfekpaphcprihceph 420
 QY 417 CKDEPSYAPYFGTNTYADTSICTAVHAGYISNESGVDVMDKKTYYGSLRNQV 476
 DB 421 ckdepsyapylgtnlyadtsicltavhagylsnesgvdvmdpdkklyvgsllrnqv 480
 QY 477 QSESLGTPRDGKAFRIEFAVRQ 497
 DB 481 qseslgtprdgkafrielfavrq 501
 RESULT 5
 AAE10617
 ID AAE10617 standard; Protein: 477 AA.
 XX
 AC AAE10617;
 DT 10-DEC-2001 (first entry)
 XX
 DE Human novel trypsin inhibitor-like protein, NOV-4c.
 XX
 KM Human: NOV-X protein: KIAA1233-like protein; STE20-like protein; tumour;
 KM trypsin inhibitor-like protein; gene therapy; haematopoietic; illness;
 KM immunological disorder; neurodegenerative disorder; Alzheimer's disease;
 KM Parkinson's disease; immunomodulatory; pharmacogenomic; haemostatic;
 KM human immunodeficiency virus; HIV; fertility disorder; neuroprotective;
 KM cytosolic; neotropic; anti-fertility; cancer; NOV-4c protein.
 OS Homo sapiens.
 XX
 PH Key Location/Qualifiers
 FT Peptide 1..22
 FT /label= Signal_peptide
 FT Protein 23..477
 FT /note= "Human mature novel trypsin inhibitor-like
 protein, NOV-4c"
 FT Domain 81..92
 FT /label= SCP_domain
 XX
 PN WO200162928-A2.
 PD 30-AUG-2001.
 XX
 PF 26-FEB-2001; 2001WO-US06151.
 XX
 PR 25-FEB-2000; 2000US-0184951.
 PR 28-FEB-2000; 2000US-0185348.
 PR 01-MAR-2000; 2000US-0185967.
 PR 18-APR-2000; 2000US-0197723.
 PR 27-APR-2000; 2000US-0199957.
 PR 23-FEB-2001; 2001US-0789390.
 XX

PA (CURA-) CURAGEN CORP.
 XX Vernet CAM, Fernandes E, Shinkets RA, Macdougall J, Spaderna SK;
 PI MPI: 2001-582051/65.
 DR N-PSDB; AAD17765.
 XX
 PT New isolated KIAA1233-like, STE20-like, or trypsin inhibitor-like
 PT polypeptide for diagnosing and treating pathological disorders, such as
 PT Parkinson's disease and for use in pharmacogenomics -
 XX
 PS Claim 1: Page 85-86; 189pp; English.
 XX
 CC The invention relates to novel human polypeptides referred as NOV-X
 CC and their corresponding nucleic acid sequences. NOV-X collectively
 CC include NOV-1, NOV-2a and NOV-2b which are novel KIAA1233-like
 CC polypeptides, NOV-3a, NOV-3b, NOV-3c and NOV-3d which are novel
 CC STE20-like polypeptides and NOV-4a, NOV-4b, NOV-4c, NOV-4d and NOV-4e
 CC which are novel trypsin inhibitor-like polypeptides. NOV-X is used to
 CC identify a potential therapeutic agent that can modulate its activity
 CC and can be used for treating a pathology related to aberrant expression
 CC or aberrant physiological interactions of NOV-X. NOV-X or its DNA is
 CC used to determine the presence or predisposition to a disease associated
 CC with altered levels of NOV-X. NOV-X, its DNA and its antibody are used
 CC to treat or prevent a pathology associated with NOV-X. The pathological
 CC states that can be treated or prevented are haematopoietic, cancer,
 CC immunological, tumour, neurodegenerative (e.g. Alzheimer's and
 CC Parkinson's disease), human immunodeficiency virus (HIV) illness and
 CC fertility disorders. NOV-X and its DNA are used in pharmacogenomics for
 CC predictive medicine. NOV-X DNA is used in gene therapy. The present
 CC sequence is human novel trypsin inhibitor-like protein, NOV-4c.
 XX
 SQ Sequence 477 AA:

Query Match 94.1%; Score 2595; DB 22; Length 477;
 Best Local Similarity 98.9%; Pred. No. 4.1e-214;
 Matches 463; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MSCVGVGVPFLGLFLFCGSGGYLLPNVTLLELLSKYOHNSHVRRAIPREDKEITL 60
 DB 1 mscvgyvfpilgllflvcgsgyllpnvlllellskymshsvrraipredkeell 60
 QY 61 MLHNKRGQVOPQASNMETWDELEKSAAMASQCIHEHPTSLVLSIGNLGAHMR 120
 DB 61 mlhnkrlrgvqgaasmeymtwdeleksaaasqcihehptslvlsignlgaahmr 120
 QY 121 YRSPGFHQSMTDEKDYTPPSCNPMCPBRCGPMCTHYTOIWMATNKGAVNTC 180
 DB 121 yrspgfhwswydevkdytppscnmpwpcrcsgpmcthytqivwatnkgavntc 180
 QY 181 RKMTVGEVMEENAVFYVNCNPKGNWIGEPYKNGRPSCEPSPSGSGCRNNLCYREET 240
 DB 181 rkmtvgevenavfyvncnpkgnwigeapkyngprpscepspsgsgcrnnlcyreety 240
 QY 241 TPKEPTDMANVEFAPLPEENHVLQPRVMPRTKPKTSANVMYQVRCDKMDRCKG 300
 DB 241 tpkpetdmevefapleehhvlqprvmrptkpktsrvnmqvlcdtkmdrckg 300
 QY 301 STCNRYCOPACCLHNHKAIFGSLYEVSSTICRAIHGIIIDDKGGLVDITRNKAVPEFV 360
 DB 301 stcnrycpgacclnhkakifgslfyessticraahygiiiddkgyglvditrnkvpfv 360
 QY 361 KSEHGVQSLSKYRPSSEFMVSKVAVDLDCTYVAQLCPPEKPAHCPRIHCPHNCDE 420
 DB 361 kserhgvqslskyrpssefmvskvkvqldcyltvaqlcppekpatihcprhpnckde 420
 QY 421 PSYAPFGCTIVADTSSICTAVHAGVISMESGDDVMPVKKKTY 468
 DB 421 psyapfgctivadtssictavhagvismesgddvmpvdkkky 468
 RESULT 6

AAE10615
 ID AAE10615 standard; Protein; 383 AA.
 XX
 AC AAE10615;
 XX
 DT 10-DEC-2001 (first entry)
 XX
 DE Human novel trypsin inhibitor-like protein, NOV-4a.
 XX
 KW Human; NOV-X protein; KIAA1233-like protein; STE20-like protein; tumour;
 KW trypsin inhibitor-like protein; gene therapy; haematopoietic; illness;
 KW immunological disorder; neurodegenerative disorder; Alzheimer's disease;
 KW Parkinson's disease; immunomodulatory; pharmacogenomic; haemostatic;
 KW human immunodeficiency virus; HIV; fertility disorder; neuroprotective;
 KW cytosolic; nontropic; anti-fertility; cancer; NOV-4a protein.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Domain 81..92
 FT /Label= SCP-domain
 XX
 PN W0200162928-A2.
 XX
 PD 30-AUG-2001.
 XX
 PF 26-FEB-2001; 2001MO-US06151.
 XX
 PR 25-FEB-2000; 2000US-0184951.
 PR 28-FEB-2000; 2000US-0185548.
 PR 01-MAR-2000; 2000US-0185967.
 PR 18-APR-2000; 2000US-0197723.
 PR 27-APR-2000; 2000US-0199957.
 PR 23-FEB-2001; 2001US-0789390.
 XX
 PA (CURA-) CURAGEN CORP.
 XX Vernet CAM, Fernandes E, Shinkets RA, Macdougall J, Spaderna SK;
 PI MPI: 2001-582051/65.
 DR N-PSDB; AAD17763.
 XX
 PT New isolated KIAA1233-like, STE20-like, or trypsin inhibitor-like
 PT polypeptide for diagnosing and treating pathological disorders, such as
 PT Parkinson's disease and for use in pharmacogenomics -
 XX
 PS Claim 1: Page 81; 189pp; English.
 CC The invention relates to novel human polypeptides referred as NOV-X
 CC and their corresponding nucleic acid sequences. NOV-X collectively
 CC include NOV-1, NOV-2a and NOV-2b which are novel KIAA1233-like
 CC polypeptides, NOV-3a, NOV-3b, NOV-3c and NOV-3d which are novel
 CC STE20-like polypeptides and NOV-4a, NOV-4b, NOV-4c, NOV-4d and NOV-4e
 CC which are novel trypsin inhibitor-like polypeptides. NOV-X is used to
 CC identify a potential therapeutic agent that can modulate its activity
 CC and can be used for treating a pathology related to aberrant expression
 CC or aberrant physiological interactions of NOV-X. NOV-X or its DNA is
 CC used to determine the presence or predisposition to a disease associated
 CC with altered levels of NOV-X. NOV-X, its DNA and its antibody are used
 CC to treat or prevent a pathology associated with NOV-X. The pathological
 CC states that can be treated or prevented are haematopoietic, cancer,
 CC immunological, tumour, neurodegenerative (e.g. Alzheimer's and
 CC Parkinson's disease), human immunodeficiency virus (HIV) illness and
 CC fertility disorders. NOV-X and its DNA are used in pharmacogenomics for
 CC predictive medicine. NOV-X DNA is used in gene therapy. The present
 CC sequence is human novel trypsin inhibitor-like protein, NOV-4a.
 XX
 SQ Sequence 383 AA:

Query Match 77.6%; Score 2140; DB 22; Length 383;
 Best Local Similarity 99.5%; Pred. No. 3.4e-175;
 Matches 379; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

SEA ID NO: 2
AC NO: AAY41738, Database: A. Geneseg-032802

OY 117 HMGRRSPGFHVQSWDEKDYTPYPSPSCNPMCPERCSCPMCTHTQIWMATTKIGCA 176
DB 3 nwgrryspbfhwsgwydevkdytppspcncpncpcscgmchtyqjwatkigca 62
OY 177 VMTGRMTWGEWENAVYFVCYSPKGNWITGAPYKNGRPPSCSPSGSCNNICR 236
DB 63 vltckmtlwgewenavyfvcnyspknwlgapyngrtpeecpssy99scnnlcyr 122
OY 237 EMTYKPEPTEDEKNEETAPRPEENHWLQPRVMPKPKRTSAVNMTQVVRCDTKMD 296
DB 123 eetyckpctedemnevetapipcenhwlpqrvmpkpkrtksaavymtgvrctkmd 182
OY 297 RCKGSTCNRYQCPAGCLNHRKATFGSLFTSSSSICRAAIHYGIIDDKGGLVDITRNKVC 356
DB 183 rckgstcnryqcpagclnhkakifgtlfyesssicraahygilddkglvdltrngkv 242
OY 357 PFFVKSERHGVOSTLSKYKSPSSFFWVSKVYODLDCYTTVAOLCFPEKPRATGPRHCPAH 416
DB 243 pffvkserrhgvostlskykspssffwvskvdyddcyltvaqlcfpekpahpripchpah 302
OY 417 CKDEPSYMAVFEGTNIYADNYSICKTAVHAGVISNESGGDVMPVDKKTYYGSLRNGV 476
DB 303 ckdepsymavfegtntiyadnysicktavhagvisnesgvdvmpvdkktyvgslrngv 362
OY 477 QSESLETPRDKAFRIRAVRQ 497
DB 363 qseslgtprdgkafiriravrq 363

RESULT 7

AAB93979 standard; Protein: 371 AA.

AC AAB93979; 26-JUN-2001 (first entry)
DB Human protein sequence SEQ ID NO:14059.
XX Human; primer: detection; diagnosis; antisense therapy; gene therapy.
XX Homo sapiens.
XX EP1074617-A2.
XX 07-FEB-2001.
XX 28-JUL-2000; 2000EP-0116126.
XX 29-JUL-1999; 99JP-0248036.
XX 27-AUG-1999; 99JP-0300253.
XX 11-JAN-2000; 2000JP-0118776.
XX 02-MAY-2000; 2000JP-0183767.
XX 09-JUN-2000; 2000JP-0241899.
XX (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI: 2001-318749/34.
XX Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX Claim 8; SEQ ID 14059; 2537bp + CD ROM; English.
XX The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary

CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination
CC of the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

Sequence 371 AA;

Query Match 75.1%; Score 2071; DB 22; Length 371;
Best Local Similarity 99.7%; Pred. No. 2.7e-169;
Matches 369; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MSCVLGVTPPLGLFLVCSQGYLIPWVTLLELLLSKYOHNSHSRVRRAIPREDKEIL 60
DB 1 mscvlgvtpplglflvcsqgyllpwtlleelllskyohnsrivrripredkeell 60
OY 61 MLHNKLRGOYQPOASNNETWMDLEKSAANAASOCITWEGPRTSLVSTQNGAHRGR 120
DB 61 mlhnklrgoyqpasnnetwmdleksaanaasocitwegprrtslvstqngahrgr 120
OY 121 YRSPGFHVQSWDEKDYTPYPSPSCNPMCPERCSCPMCTHTQIWMATTKIGCAVNTC 180
DB 121 yrspgfhwswdevkdytppspcncpncpcscgmchtyqjwatkigcavntc 180
OY 181 RKMTWGEWENAVYFVCNTPKGNWITGAPYKNGRPPSCSPSGSCNNICRETY 240
DB 181 rkmtlwgewenavyfvcnyspknwlgapyngrtpeecpssy99scnnlcrety 240
OY 241 TPKPEDEKNEETAPRPEENHWLQPRVMPKPKRTSAVNMTQVVRCDTKMDKRCG 300
DB 241 tpkpetdemnevetapipcenhwlpqrvmpkpkrtksaavymtgvrctdkmkrckg 300
OY 301 STCNRYQCPAGCLNHRKATFGSLFTSSSSICRAAIHYGIIDDKGGLVDITRNKVC 360
DB 301 stcnryqcpagclnhkakifgtlfyesssicraahygilddkglvdltrngkvpfv 360
OY 361 KSERHGVOSTLSKY 370
DB 361 kserhgvostlsky 370

RESULT 8

AAV41738 standard; Protein: 500 AA.

AC AAV41738;
XX 07-DEC-1999 (first entry)
XX Human PRO541 protein sequence.
XX Human; PRO: EST: expressed sequence tag; PCR primer; hybridisation;
XX probe; blood coagulation disorder; cancer; cellular adhesion disorder;
XX secreted protein; transmembrane protein.
XX Homo sapiens.
XX OS
XX PN W09946281-A2.

XX	16-SEP-1999.	99MO-0505028
PD		
XX	08-MAR-1999.	
PF		
XX		
PR	10-MAR-1998.	98US-0077450
PR	11-MAR-1998.	98US-0077632
PR	11-MAR-1998.	98US-0077641
PR	11-MAR-1998.	98US-0077659
PR	12-MAR-1998.	98US-007791
PR	13-MAR-1998.	98US-0078004
PR	17-MAR-1998.	98US-0040220
PR	20-MAR-1998.	98US-0078886
PR	20-MAR-1998.	98US-0078910
PR	20-MAR-1998.	98US-0078939
PR	25-MAR-1998.	98US-0079839
PR	26-MAR-1998.	98US-0079294
PR	27-MAR-1998.	98US-0079656
PR	27-MAR-1998.	98US-0079663
PR	27-MAR-1998.	98US-0079664
PR	27-MAR-1998.	98US-0079689
PR	27-MAR-1998.	98US-0079728
PR	30-MAR-1998.	98US-0079786
PR	30-MAR-1998.	98US-0079920
PR	31-MAR-1998.	98US-0079923
PR	31-MAR-1998.	98US-0080105
PR	31-MAR-1998.	98US-0080107
PR	31-MAR-1998.	98US-0080165
PR	31-MAR-1998.	98US-0080194
PR	01-APR-1998.	98US-0080327
PR	01-APR-1998.	98US-0080328
PR	01-APR-1998.	98US-0080333
PR	01-APR-1998.	98US-0080334
PR	08-APR-1998.	98US-0081079
PR	08-APR-1998.	98US-0081070
PR	08-APR-1998.	98US-0081071
PR	09-APR-1998.	98US-0081103
PR	09-APR-1998.	98US-0081293
PR	09-APR-1998.	98US-0081829
PR	15-APR-1998.	98US-0081817
PR	15-APR-1998.	98US-0081838
PR	15-APR-1998.	98US-0081952
PR	21-APR-1998.	98US-0081955
PR	21-APR-1998.	98US-0082356
PR	21-APR-1998.	98US-0082569
PR	22-APR-1998.	98US-0082704
PR	22-APR-1998.	98US-0082704
PR	23-APR-1998.	98US-0082767
PR	23-APR-1998.	98US-0082796
PR	28-APR-1998.	98US-0083336
PR	29-APR-1998.	98US-0083322
PR	29-APR-1998.	98US-0083495
PR	29-APR-1998.	98US-0083496
PR	29-APR-1998.	98US-0083500
PR	29-APR-1998.	98US-0083545
PR	29-APR-1998.	98US-0083554
PR	29-APR-1998.	98US-0083558
PR	30-APR-1998.	98US-0083559
PR	30-APR-1998.	98US-0083742
PR	06-MAY-1998.	98US-0084366
PR	06-MAY-1998.	98US-0084414
PR	07-MAY-1998.	98US-0084441
PR	07-MAY-1998.	98US-0084598
PR	07-MAY-1998.	98US-0084600
PR	07-MAY-1998.	98US-0084627
PR	07-MAY-1998.	98US-0084637
PR	07-MAY-1998.	98US-0084639
PR	07-MAY-1998.	98US-0084640
PR	13-MAY-1998.	98US-0084643
PR	13-MAY-1998.	98US-0085323

Wood WL, Goddard A, Gurney A, Yuan J, Baker KP, Chen J;
WPI: 1999-551358/46.
N-PSDB: AAZ34199.

New secreted and transmembrane polypeptides and their polynucleotides
useful for treating blood coagulation disorders, cancers and cellular
adhesion disorders -

Claim 12, Fig 147, 530pp; English.

The present invention describes secreted and transmembrane polypeptides and their polynucleotides. The nucleotide sequences are useful as sources of probes, primers, for chromosome mapping, and for generation of antisense sequences. They can also be used to create transgenic animals. The proteins can be used to treat a variety of diseases and disorders, depending on their function. Diseases that may be treated include blood coagulation disorders, cancers and cellular adhesion disorders. They may also be used to raise antibodies. AAZ33891 to AAZ34338, and AA14085 to AA141774 represent polynucleotides and polypeptide sequence given in the exemplification of the present invention.

Sequence 500 AA;
xx

Query Match	59.5%;	Score 1640.5;	DB 20;	Length 500;
Best Local Similarity	59.5%;	Pred. No. 3.7e-132;		
Matches 292;	Conservative 74;	Mismatches 112;	Indels 13;	Gaps 7

```

QY 13 LLEFWGSGQGLPVPVLTLEELSTGYONE-----SHSRNRRAIPREDKEETIMLHANKLR 67
Db 14 vlfmaiaipamvvpnatllleklliekymdegewwjakqrgfartldndmgslldlhnklir
QY 68 GOVOPASNMETMTWDDLEKSAAMAASQCITWEGPRTSLVSIQNLGAIHGRSPGFH 127
Db 74 sqvyplassmewmtcdvelersaeswaesciwehgpsallpslsqnlghahgryrptfth 133
QY 128 VQSNVDEYKDDTYTYPRSCNCPWCEBRCGSPMCTHYTQIWAITTKIGCAVNTCKRMTWG 187
Db 134 vqswydeydkdtsyehcncpfcfrcsfspacthytyqvwatnrrlgaiahlnchmmiawg 193
QY 188 EWMENAVYWCNISPKNMIGEALEYKKNGRPCSECPSPGSGCRNLTYRE--ETTPPKPE 245
Db 194 qlwpykavylwcnyspknugwshaprykngfprsaacprrfsggcrcenllyksgdstypr-pr 252
QY 246 TDEENVEETAPIR-EEHNHYMLQPRVMPRTYPRKKTSAVNVMYMTQVVRCDTKMKKDRCKSTCN 304
Db 253 eeeenleierqsgvqhchlv--rttsddasanevysa-qmgsqslfscvzrlldcokctcn 309

```

```
QY 305 RYOCPCACGLNKKAKIFGSLFYESSSSICRAAIHYGILDDKGLVDITRNGKVPFVKSR 364
DB 310 rycpcagclskakvlgsvhyemgssicraahygildnggwldlrggrkhyfiksnt 369
QY 365 HGVQSLSKYKPSFMSKVVQDLDCYTYAQLCPFEKPAHCRHCPACCKDEPSYW 424
DB 370 ngdltgkysansftvskvgtvctveqlcpfhkpsahcpryvcprcmqanphy 429
QY 425 APVFGTNIYADTSSICRAAHAGVINSNCGDVPDVPDKKTYGSLRNGVQSSSLGTP 484
DB 430 arvlgtrvysldssicraahagvvrnh-ggyvdvmpdkrtktylasfnglfseslgnp 488
QY 485 RDGKAFRIFAV 495
DB 489 pggkaftrfav 499

RESULT 9
ID AAB44294 standard; Protein: 500 AA.
AAB44294;
AC AAB44294;
DT 08-FEB-2001 (first entry)
DE Human PRO541 (UNQ342) protein sequence SEQ ID NO:363.
XX Human: secreted protein; transmembrane protein; PRO; EST; cytosolic;
KW expressed sequence tag; detection; cancer.
OS Homo sapiens.
XX
XX WO200053756-A2.
PD 14-SEP-2000.
XX
XX 18-FEB-2000; 2000WO-US04341.
XX
XX 08-MAR-1999; 99WO-US05028.
XX 12-MAR-1999; 99US-0123957.
XX 29-MAR-1999; 99US-0126773.
XX 21-APR-1999; 99US-0130232.
XX 28-APR-1999; 99US-0131445.
XX 14-MAY-1999; 99US-0134287.
XX 23-JUN-1999; 99US-0141037.
XX 26-JUL-1999; 99US-0145698.
XX 29-OCT-1999; 99US-0162506.
XX 30-NOV-1999; 99WO-US28313.
XX 02-DEC-1999; 99WO-US28551.
XX 02-DEC-1999; 99WO-US28565.
XX 16-DEC-1999; 99WO-US30095.
XX 30-DEC-1999; 99WO-US31243.
XX 30-DEC-1999; 99WO-US31274.
XX 05-JAN-2000; 2000WO-US00219.
XX 06-JAN-2000; 2000WO-US00277.
XX 06-JAN-2000; 2000WO-US00376.
XX
XX (GETH ) GENENTECH INC.
XX
XX Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;
XX Ferrara N, Filvaroff E, Fong S, Gao W, Garber H, Gerlitsen ME;
XX Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;
XX Kljavin IJ, Kuo SS, Napier MA, Pan J, Paoni NF, Roy MA;
XX Shelton DL, Stewart TA, Tumas D, Williams PM, Wood WI;
XX
XX WPI: 2000-611443/58.
XX N-PSDB: AAC78550.
XX
XX Novel PRO polypeptides and polynucleotides used in detection methods,
XX to target bioactive molecules to specific cells, and to modulate
XX cellular activities -
XX
XX Claim 12; Fig 147; 636pp; English.
```

```
XX AAC78458 to AAC78599 represent polynucleotide and EST (expressed
CC sequence tag) sequences which encode secreted or transmembrane PRO
CC polypeptides. The PRO polynucleotides and polypeptides have cytosolic
CC activity. The polynucleotides and polypeptides can be used for detecting
CC the presence of PRO polypeptides in samples, for linking bioactive
CC molecules to cells and for modulating biological activities of cells,
CC using the polypeptides for specific targeting. The polypeptide targeting
CC can be used to kill the target cells, e.g. for the treatment of cancers.
CC The polypeptide pairs provide specific targeting of bioactive molecules
CC to cells. AAC78600 to AAC78987 represent PCR primers and probes used in
CC the isolation of the PRO polynucleotide sequences.
XX
XX Sequence 500 AA:
SQ
Query Match 59.58; Score 1640.5; DB 21; Length 500;
Best Local Similarity 59.58; Pred. No. 3.7e-132;
Matches 292; Conservative 74; Mismatches 112; Indels 13; Gaps 7;
QY 13 LFLVCGSGCYLBNVTLLEELSKYQHN-----SHSRVRAIPREDKEILMLNKLK 67
DB 14 vlfmarapamvvpnatlleklekyndedegwylakrgkraltndmgslldlnklr 73
QY 68 GQVQPOASNMEYMTWDELKESAAAMASQCTWHEGPTSLVSGNLCAMHGRYSPGFH 127
DB 74 sqvyltasmeymtwdelersaeswescilwehgpasllpslqnlghwgyryypctfh 133
QY 128 VQSYVDEVKDVTYRPSRCNWPCEKSGPCMTHTYTOIWMATTNKIGCAVNTCRKMTWG 187
DB 134 vqswydevkdstyphecnpcfcscgpccthyqvwatnrlgcalnchmmlwg 193
QY 188 EWMENAVYVCNYSNPKNGWIGAPYKNGRPSCEPSSYSGSRRNLGYRE--EYTPPE 245
DB 194 qlpwkavylvncypkgwmgahykhgrpsacpsfsggrenlcyksgsdrlyp-pr 252
QY 246 TDEMNVEYETAPR-ENNVMYQPRMRTKPKTSAAVVMYQVRCMDKRCGSGTCN 304
DB 253 eectneletrqsgvhdttv--rttsdssrnevsa-qgmsqlvseevrlltdqcgylton 309
QY 305 RYOCPCACGLNKKAKIFGSLFYESSSSICRAAIHYGILDDKGLVDITRNGKVPFVKSR 364
DB 310 rycpcagclskakvlgsvhyemgssicraahygildnggwldlrggrkhyfiksnt 369
QY 365 HGVQSLSKYKPSFMSKVVQDLDCYTYAQLCPFEKPAHCRHCPACCKDEPSYW 424
DB 370 ngdltgkysansftvskvgtvctveqlcpfhkpsahcpryvcprcmqanphy 429
QY 425 APVFGTNIYADTSSICRAAHAGVINSNCGDVPDVPDKKTYGSLRNGVQSSSLGTP 484
DB 430 arvlgtrvysldssicraahagvvrnh-ggyvdvmpdkrtktylasfnglfseslgnp 488
QY 485 RDGKAFRIFAV 495
DB 489 pggkaftrfav 499

RESULT 10
ID AAU29058 standard; Protein: 500 AA.
AAU29058;
AC AAU29058;
DT 18-DEC-2001 (first entry)
DE Human PRO polypeptide sequence #35.
XX
XX PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep;
XX dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;
XX blood; chondrocyte cell; cell proliferation; cell differentiation; colon;
XX adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder.
XX
XX Homo sapiens.
OS
```


QY 121 YRSPGFHVQSWYDEKDYTPYSECNPNWCPERCSPWCTHYTOIWAATTKIGCAVNTC 180
 |||
 Db 121 YRSPGFHVQSWYDEKDYTPYSECNPNWCPERCSPWCTHYTOIWAATTKIGCAVNTC 180
 QY 181 RKTATWGEWENAVYFCVNSPKGNWIGEAEPYKNGRCPSECPSPYSGSCRNNLCYREPT 240
 |||
 Db 181 RKTATWGEWENAVYFCVNSPKGNWIGEAEPYKNGRCPSECPSPYSGSCRNNLCYREPT 240
 QY 241 T 241
 |||
 Db 241 T 241
 |||

RESULT 13

AAB41816
 ID AAB41816 standard; Protein; 431 AA.

AC AAB41816;

DT 08-FEB-2001 (first entry)

DE Human ORFX ORF1580 polypeptide sequence SEQ ID NO:3160.

Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 vulnery; antipsoriatic; antiparkinsonian; neurotropic; neuroprotective;
 anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac;
 hypotensive; dermatological; immunosuppressive; antidiabetic;
 antiviral; antibacterial; antifungal; antirheumatic; antihypertensive;
 antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 neurodegenerative disorder; osteoarthritis; graft vs host disease;
 cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 cholesterol ester storage; systemic lupus erythematosus; infection;
 severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 allergy; aplastic anemia; nocturnal haemoglobinuria; burn; wound;
 bone damage; cartilage damage; antiinflammatory disease; coagulation;
 thrombosis; contraceptive.

OS Homo sapiens.

PN WO200058473-A2.

PD 05-OCT-2000.

PF 31-MAR-2000; 2000WO-US08621.

PR 31-MAR-1999; 99US-0127607.

PR 02-APR-1999; 99US-0127636.

PR 05-APR-1999; 99US-0127728.

PR 30-MAR-2000; 2000US-0540763.

PA (CURA-) CURAGEN CORP.

PI Shinkets RA, Leach M;

DR WPI: 2000-602362/57.

DR N-PSDB; AAC76025.

PS Claim 11; Page 2377-2378; 5507pp; English.

XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,

CC which represent the human ORFX open reading frames 1 to 3161. The ORFX

CC sequences have activities such as: cytostatic; hepatotropic; vulnery;

CC antipsoriatic; antiparkinsonian; neurotropic; neuroprotective;

CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;

CC immunostimulant; cardiac; thrombolytic; coagulant; vasotropic;

CC antidiabetic; hypotensive; dermatological; immunosuppressive;

CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;

CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 CC
 XX Sequence 431 AA;
 SQ
 Query Match 47.28; Score 1302; DB 21; Length 431;
 Best Local Similarity 56.48; Pred. No. 3.4e-103;
 Matches 237; Conservative 61; Mismatches 96; Indels 26; Gaps 8;
 QY 13 LLEFVCGSGYLLPNTLLEELSKYQNE-----SHSRVRAIPREDKEEIMLHAKLR 67
 |||
 Db 14 VLMARALPAMVVPNALLLEKLEKYMDEGEWIAKRGKRALTDMDGSLDHLKLR 73
 QY 68 GQVQPOASNNEMWTWDELKSAAMASOCIMERGPTSLVTSQNTGAHGRYRSPGFH 127
 |||
 Db 74 SQVPLASMEYMTWDELERSAESWESCLWEHGPSLLPSISQNLGAHWGRYRPTFH 133
 QY 128 VQSYVDKVKYTYTPYSECNPNWCPERCSPWCTHYTOIWAATTKIGCAVNTCKRMVWG 187
 |||
 Db 134 VQSYVDKVKYTYTPYSECNPNWCPERCSPWCTHYTOIWAATTKIGCAVNTCKRMVWG 193
 QY 188 EWEENAVYFCVNSPKGNWIGEAEPYKNGRCPSECPSPYSGSCRNNLCYRE- EYTPKPE 245
 |||
 Db 194 GWPKAYVLYCNSPKGNWIGEAEPYKNGRCPSECPSPYSGSCRNNLCYRE- EYTPKPE 252
 QY 246 TDEKNEVETAPIP- EENHVMLOPRVMPYKPKTSAAVNTQVVRCDTKMKDRCKSTCN 304
 |||
 Db 253 EEEENELERQSGVHDTLV--RTSDSSRNEVISA-qmnsqjvscvrlrdqgkytcn 309
 QY 305 RYQCPAGCLNHNKAKIFGSLFPESSSICRAAHYGIIDMGGLVDITRNKVPFVNSER 364
 |||
 Db 310 RYQCPAGCLNHNKAKIFGSLFPESSSICRAAHYGIIDMGGLVDITRNKVPFVNSER 369
 QY 365 HGVOQLSKYKPPSSFFMSKVKVODLDCYTYV-----AQCPFKPATHCRIH 412
 |||
 Db 370 NGIQTGKYSAANSFCVSKVTVGAVICETLVDSVHISLHLAG--EYLVLTVCQGLN 427
 RESULT 14
 AAM39907
 ID AAM39907 standard; Protein; 300 AA.
 AC AAM39907;
 DT 22-OCT-2001 (first entry)
 DE Human polypeptide SEQ ID NO 3052.
 XX Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.
 XX Homo sapiens.
 XX WO200153312-A1.
 XX 26-JUL-2001.
 PD 26-DEC-2000; 2000WO-US4263.

CC Reis-Bucklers corneal dystrophy. This sequence represents FCTR7, a
CC trypsin inhibitor-like protein.
CC
xy

